

FT	REPEAT	2778	2781	6.	
FT	REPEAT	2782	2785	7.	
SQ	SEQUENCE	2867	AA;	331433	MW;
	Query Match		4.3%;	Score 231;	DB 1;
	Best Local Similarity		18.9%;	Pred. No. 0.0055;	Length 2867;
Matches	243;	Conservative	219;	Mismatches	436;
				Indels	388;
				Gaps	65;
QY	11	LALITTVWVTVSQEWYGLEREESVKQEQTOS---	ASEDDWFEEDNERKTNVSKENSTV-	65	
Db	1314	IATTKKGTSPTSDINELL--ESIKEEVHKNLQLOVKQESNSMEE--MRQOISWMKDLITL	1368		
QY	66	--DETUSDLFSDG-----NSNNSSKTESVVS-----DPKQVPKAKPVEVQFQASH	108		
Db	1369	NNSETIAKEIISNNNTQNAIGFRENAAKTLINKTDELLQRAAMIEEAKAHNNIDIALEDAQ	1428		
QY	109	SSNDASKWVVPKQDTASKETLETSTWRAKDFVTRGDTLVGFSKSGINKLQSQTSHLVPS	168		
Db	1429	IDTEVSKTEQINREIMNKKDETISKYLSEIKEYKDKCTTEISNSKRKGDKIEFLE-----	1482		
QY	169	HAADGTQLTQVASPAFTPDK---KTAIAEYTSRLGENGKPSRDLID--QKEHILDEGET	221		
Db	1483	-----KEFPNEESENNSNKVNININEINENI--RNSEQYIQLDIEAEKQASTKVEL	1527		
QY	222	FNAYQVLTQKLTIPNGYKSIGQDAFDVNKNIAEVNLPELETISDVAFAHMSLKVQKLPDNL	281		
Db	1528	FHKHE--TTISNIKE-SEILGVETKSQOKINKAEDIMKEIERHNEIQTQVKGFQENTL	1583		
QY	282	KVIGELAFFFNDQIGGKXKMLPRHLIKLAERAFAKSNSRIOQTVEFLGSKLKVIGEASFOD--NN	339		
Db	1584	INKLNPHNYDN-----AEDELNNDK-----STNAKVLIENTLESVKHN	1621		
QY	340	LRNVM-LPDGLEKIES-----EAFTGNPGDE-----HYNNOVVLRTT	376		
Db	1622	ISEITNIKOGGEKIYSKAKDIMOKIKATSENATAEKTLERVKDDQSNVNVNLQIT---T	1677		
QY	377	GQNPHQHOLATE-----NTYVNPDKSL--WRATPDMDYTKWLE-----DFTYQK	417		
Db	1678	ERN--LIVTEKNRNLGIDSTTINTEGALKESKGNYEIGFPLEKLEIGKNRKLKVDTIKKS	1735		
QY	418	-NSVTG-----FSNKGLQKVRNNKL-----EIPKQHNGITTEIGDNAFRNV---	459		
Db	1736	INSTVGNFSSLNNFDINQYDFNKNINDYENKRMGEIYNEFEG-SLNKISEN-LRNASENT	1793		
QY	460	-DFQOS-KTLR-----KYDL----EETKLPSTIRKIGAFAQFQSNNLKSFEASEDLEI-	505		
Db	1794	SDYNSAKTLRLEAQEKVNLLNKEEANKYLDRDVVKVESFRF-----IFNMKESLDKIN	1847		
QY	506	-----KEGAFMNRIGHTL-DLKDKLLIKIGDAAFHINHITYAIVLP-----ESVQEIGRSAFR	555		
Db	1848	EMIKKEQLTVNNEGHGNVKQVLVENNIKELVDE---NNLSDILKQATGKNEEIQKITHSTLK	1903		
QY	556	ONGALHLMFIGNKVKTIG-----EMAFLSNKLESVNLSSEQKQLK-----TIEVQAFSD	603		
Db	1904	NKAKTILGHVHTSARKVGKIKITPELA-LTELLGDAKLTQALKEFESKNNTVLETENMSK	1962		
QY	604	NALSEVVLPPNLQ-----TIREEAFK-----RNHL	628		
Db	1963	NT-NELDVHKNIQDAYKVALEILAHSDDEIDTKQKDSSKLIEMGNQIYLKVVLINQYKNKI	2021		
QY	629	KEVKGS-----SILSQTIE-----HNNSHM-----NADQ-----NAFDQ-----N	648		
Db	2022	SSIKSKEEAVSVKIGNVSKHSELSKITCSDSKSYDNTIALEKQTELQNLRNNSFTQEKNT	2081		
QY	649	DGDKRGK-----KVVRT-----HNNSHM-----LAGDERFIID	678		
Db	2082	NSDSKLEKIKTDFESLKNALKTLEGEVNAALKASSDNHBEHQSKSEPVNPALSEIEKEED	2141		
QY	679	PDKLSSTMVDILEKVKLIEGLDYSTLROTTQFREMTA-----GKALLS--KSNLR	729		
Db	2142	IDSINTALDELLKKGRCTEVSRYKLKDVTKEISDDTELINTIEKNVKAYAVIKNYE	2201		
QY	730	QGEKQKFLQBAQFFLGRV-----DLDKATAKAELVTKKATKNGHLLERSINKAVLX	783		

Db	2202	DTWQDVLTLINEHFNTKQVSNEHTNFDKSNSSEE--LTXAVTDSKTIISKL--KGVIIE	2257	
QY	784	NNSAIKKANVKRLEKEELDLTDLVECKGGLAQMATTWQCVYLLKTPPLPEYYIGLVNVYFD	843	
Db	2258	VNENTEMNTIESAKAIEALYNELNKK-----KTSNEIYVTSNEVKLOEMKSNAADKYID	2312	
QY	844	KSGKLIYALDMSDTIGEOKDAYGMPILNVNDEBNRGYHTLAVATLADYEGLYIKDILNS	903	
Db	2313	VS-----KIENTVLDTQK--SNIVTNOHSINN-----VKDKLKGK	2345	
QY	904	LDK-IKAIRQIPLAKYHRLG-IFQAIRNAAEADRLLPKTPKGYLNEVNPYRKQOMEKNL	961	
Db	2346	LQELIDADSSTLESIKFNEIYSHKTINIGELEQ-LQQTNKSEHDNVAKHEKTIH--	2401	
QY	962	KPVDYKTPFLNPKNEVKGDRAKH-----NINAETTNSVAVTPIRSE	1007	
Db	2402	--LINRV--ESLKGD--VKNHDDQYQMKKLNASLINDNKTNTNSINISDEELK	2448	
QY	1008	QQLHKSQDVNLQPTSSKNNFYIEL	1033	
Db	2449	KLIKKEVENDQCLICKNNNTQFISD	2474	
RESULT 2				
ID	NUM1	YEAST	STANDARD;	PRT;
AC	00402	Q03767;		2748 AA.
DT	01-DEC-1992	(Rel. 24, Created)		
DT	10-OCT-2003	(Rel. 42, Last sequence update)		
DE	Nuclear migration protein	NUM1		
GN	NUM1 OR YDR150W OR YD8358-06.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OX	Saccharomyces cerevisiae; Saccharomycetaceae; Saccharomyces; NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 28383 / FL100;			
RC	MEDLINE=97313263; PubMed=9169867;			
RA	Jacq C., Alt-Moerle J., Andre B., Arnold W., Bahr A., Ballesta J.P.G.,			
RA	Barbosa M., Baron L., Becker A., Biteau N., Bloecker H., Blugeon C.,			
RA	Boskovic J., Brandt P., Brueckner M., Buitrago M.J., Coster F.,			
RA	Bargues M., Baron L., Becker A., Biteau N., Bloecker H., Blugeon C.,			
RA	Delaveau T., del Rey F., Dujon B., Eide L.G., Garcia-Cantalejo J.M.,			
RA	Goffeau A., Gomez-Perez A., Granotier C., Hanemann V., Hankeln T.,			
RA	Hoheisel J.D., Jaeger W., Jimenez A., Jorniau J.-L., Kraemer C.,			
RA	Kuester H., Laamanen P., Legros Y., Louis E.J., Moeller-Rieker S.,			
RA	Monnet A., Moro M., Mueller-Auer S., Nussbaumer B., Paricio N.,			
RA	Paulin L., Perez J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,			
RA	Prydz H., Purnelle B., Rasmussen S.W., Remacha M., Revuelta J.L.,			
RA	Rieger M., Salom D., Saluz H.P., Saiz J.E., Saren A.-M., Schaefer M.,			
RA	Scharfe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,			
RA	Urrestarazu L.A., Verhasselt P., Vissers S., Voet M., Volckaert G.,			
RA	Wagner G., Wambutt R., Wedder E., Wedder H., Woelfl S., Harris D.E.,			
RA	Bowman S., Brown D., Churcher C.M., Connor R., Dedman K., Gentles S.,			
RA	Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,			
RA	Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,			
RA	Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen B.,			
RA	Araujo R., Aviles E., Berino A., Carpenter J., Chen E., Cherry J.M.,			
RA	Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,			
RA	Lashkari D., Lew H., Lin D., Mosedale D., Nakahara K., Namath A.,			
RA	Oefner P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,			
RA	Shogren T., Shroff N., Wiant A., Yelton M.A., Botstein D.,			

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 27, 2004, 11:10:26 ; Search time 17 seconds
 (without alignments)
 3231.413 Million cell updates/sec

Perfect score: US-10-091-007A-24
 Sequence: 5352
 1 MTKKHLKTLALALTTVSVT.....VSLCLLFLVTAGKKRKARK 1055

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	231	4.3	2867	1	RBP2_PLAVB	Q00799 plasmidium
2	224	4.2	2748	1	NUM1_YEAST	Q00402 saccharomyce
3	215.5	4.0	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
4	209	3.9	1208	1	PCP1_SCHPO	Q92351 schizosacch
5	209	3.9	1875	1	MLP1_YEAST	Q02455 saccharomyce
6	208	3.9	1013	1	SCA4_RICRH	Q9aj81 rickettsia
7	203	3.8	998	1	SCA4_RICAK	Q9ajx9 rickettsia
8	202.5	3.8	2869	1	RBP1_PLAVB	Q00798 plasmidium
9	202	3.8	1022	1	SCA4_RICPR	Q9zd49 rickettsia
10	200	3.7	872	1	TE95_TETTH	Q94819 tetrahymena
11	200	3.7	939	1	IF2_WOLSU	Q7m7x5 wolinella s
12	199.5	3.7	3259	1	GOBT_HUMAN	Q14789 homo sapien
13	199	3.7	1357	1	KTN1_HUMAN	Q86up2 homo sapien
14	199	3.7	1957	1	SP0F_SCHPO	Q10411 schizosacch
15	198.5	3.7	1790	1	US01_YEAST	Q25386 baccharomyce
16	198	3.7	1010	1	SCA4_RICPA	Q9aj75 rickettsia
17	196.5	3.7	992	1	SCA4_RICRY	Q9aj63 rickettsia
18	195.5	3.7	1969	1	MYSA_CAEEL	P12844 caenorhabdi
19	192.5	3.6	1358	1	SIR4_YEAST	P11978 saccharomyce
20	192.5	3.6	1839	1	CYAA_SACKL	P23466 baccharomyce
21	192	3.6	1022	1	SCA4_RICCN	Q52658 rickettsia
22	189.5	3.5	981	1	SCA4_RICFE	Q9aj37 rickettsia
23	188.5	3.5	1012	1	SCA4_RICSL	Q9aj80 rickettsia
24	188	3.5	1011	1	SCA4_RICAF	Q9aj83 rickettsia
25	187	3.5	1928	1	MYSI_YEAST	P08964 saccharomyce
26	186.5	3.5	1011	1	SCA4_RICMN	Q9aj82 rickettsia
27	185	3.5	1018	1	SCA4_RICJA	Q9aj79 rickettsia
28	185	3.5	1972	1	MYHB_RABIT	P35748 oryctolagus
29	184.5	3.4	1539	1	Y373_HUMAN	Q15078 homo sapien
30	183.5	3.4	1315	1	CHAO_DROME	P12024 drosophila
31	183.5	3.4	8545	1	ANCI_CAEEL	Q9n4m4 caenorhabdi
32	183	3.4	2663	1	CENE_HUMAN	Q02224 homo sapien
33	3.4	991	1	SCA4_RICSI	Q9aj77 rickettsia	

ALIGNMENTS

RA Davis R.W., Johnston M., Andrews S., Brinkman R., Cooper J., Ding H.,
RA Du Z., Favello A., Fulton L., Gattung S., Greco T., Hallsworth K.,
RA Hawkins J., Hillier L.W., Jier M., Johnson D., Johnston L.,
RA Kirsten J., Kucaba T., Langston Y., Latreille P., Le T., Mardis E.,
RA Menezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin L.,
RA Riles L., Taich A., Trevaskis E., Vignati D., Wilcox L., Woldman P.,
RA Vaudin M., Wilson R., Waterston R., Albermann K., Hani J., Heumann K.,
RA Kleine K., Mewes H.-W., Zollner A., Zaccaria P.;
RT "the nucleotide sequence of *Saccharomyces cerevisiae* chromosome IV.";
RL *Nature* 387:75-78 (1997).

-!- FUNCTION: Controls nuclear migration. NUM1 specifically controls the interaction of the bud neck cytoskeleton with the pre-divisional G2 nucleus perhaps by recognizing G2-specific cytoplasmic microtubuli or other components of the nuclear envelope.

-!- MISCELLANEOUS: Additional regions of lower homology to the repeat consensus (always starting with proline) are found in both flanking domains of the tandem repeats.

-!- SIMILARITY: Contains 1 PH domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC 178 QVASFAFTPDKKTAIAEYTSRIGKPSRLDIDQETIDEGEIFNAYQLTKLITPQYK 237
CC 304 DPDSEGTISSEEDIFDIVIEIDMIEITGSPVEDISIEDLVKKYSEKNMILSNDSYKLNQ 363
CC 364 KSESASKPKDDELMTKEVAENLNMTALPNDNYSKKEFSLESHI-----KYLEASGYK 416

CC 238 SIGQDAFVDNKNTAENV----LPESLETI-----
CC 417 VLPLEEF-ENLNESLSNPSYNTVKEKLOAKKIPIDQSTFNLKEPTIDFLPLTSKIDC 475

CC 263 ----SDY----AFAHMSLKQVK-----LPDNL-----KVIGELAFFD--N 292
CC 476 LIITPKDYNLFEVKNSPKSIEQMKKCLEAKNDLQSNICKWLEERNGCKWLSDLYFSMVN 535

CC 293 QI--GGKLYLPRHL----IKLAERAFKSNRIOTVEFLGSKLVIGEASFQDNLL--RNV 343
CC 536 KIETPSKQYLSDAKEKEYDQVLIDTKALEGLKNPTIDFLREK----ASASDYLILKKED 589

CC 344 MLPDGLEKTESEAFTGNP---GDEHYNQV-----VLRTRTGQNPOLATENTYVN 391
CC 590 YVSPSLEYLVEHAKATNHLILSDSAYEDLVCKKENPDMEFLKEKSAKLGHTVSNEAYSE 649

DR EMBL; Z50046; CAA43554.1; --.
DR GermOnline; 140641; --.
DR SGD; S0002557; NUM1.

DR GO; GO:0005934; C:bud tip; IDA.
DR GO; GO:0005938; C:cell cortex; IDA.
DR GO; GO:0015631; F:tubulin binding; IPI.
DR GO; GO:0000226; P:microtubule cytoskeleton organization and b. . . ; IMP.
DR GO:000065; P:nuclear migration (sensu *Saccharomyces*); IMP.

DR InterPro; IPR05529; DUF321.
DR InterPro; IPR01849; PH.
DR Pfam; PF03778; DUF321; 13.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS5003; PH_DOMAIN; 1.

KW Repeat.

FT DOMAIN 593 1384 12.5 X TANDEM REPEATS.

FT REPEAT 593 656 1.
FT REPEAT 657 727 2.
FT REPEAT 728 798 3.
FT REPEAT 799 862 4.
FT REPEAT 863 926 5.
FT REPEAT 927 990 6.
FT REPEAT 991 1054 7.
FT REPEAT 1055 1118 8.
FT REPEAT 1119 1182 9.
FT REPEAT 1183 1246 10.
FT REPEAT 1247 1310 11.
FT REPEAT 1311 1374 12.
FT REPEAT 1375 1384 13. (INCOMPLETE).

FT DOMAIN 2573 2683 PH.

FT CONFLICT 1570 1570 A -> V (IN REF. 1).
FT CONFLICT 1822 1822 E -> K (IN REF. 1).
FT CONFLICT 1960 1962 KAS -> RHL (IN REF. 1).
FT CONFLICT 1971 1972 KQ -> RN (IN REF. 1).
FT CONFLICT 2049 2049 S -> N (IN REF. 1).
FT CONFLICT 2637 2637 V -> A (IN REF. 1).
SQ SEQUENCE 2748 AA; 313030 MW; EB4E48F950621142 CRC64;

Query Match 4.2%; Score 224; DB 1; Length 2748;
Best Local Similarity 20.0%; Pred. No. 0.011; Mismatches 228; Conservative 191; Indels 258; Gaps 50;

QY 7 KTLALALTIVSVVTYSOEVYGLEREESVKQEQTQSASEDDWFEEBNERKTNVSKENISTVD 66
Db 1174 SELEKKLEOPLSIALIVE--HAKATDHHTSDSAYEDLVCKKENPDVEFLK-----EKSALKLIGHTVSNAY 1173

QY 67 ETVSDLFSDGNSSNNSSKTESVVSDPKQVKAKPFTQ-----EASNSSNDASKVE 117
Db 246 ISVSDLTSEINS-LQSNTER-IKIQQDPAKASSILKRVQKYYQKOHTSDTVS 303
Db 118 VPKQDTASKKETLETTSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSHAADGTQLT 177
Db 304 DPDSEGTISSEEDIFDIVIEIDMIEITGSPVEDISIEDLVKKYSEKNMILSNDSYKLNQ 363

QY 238 SIGQDAFVDNKNTAENV----LPESLETI-----
Db 417 VLPLEEF-ENLNESLSNPSYNTVKEKLOAKKIPIDQSTFNLKEPTIDFLPLTSKIDC 475

QY 263 ----SDY----AFAHMSLKQVK-----LPDNL-----KVIGELAFFD--N 292
Db 476 LIITPKDYNLFEVKNSPKSIEQMKKCLEAKNDLQSNICKWLEERNGCKWLSDLYFSMVN 535

QY 293 QI--GGKLYLPRHL----IKLAERAFKSNRIOTVEFLGSKLVIGEASFQDNLL--RNV 343
Db 536 KIETPSKQYLSDAKEKEYDQVLIDTKALEGLKNPTIDFLREK----ASASDYLILKKED 589

QY 344 MLPDGLEKTESEAFTGNP---GDEHYNQV-----VLRTRTGQNPOLATENTYVN 391
Db 590 YVSPSLEYLVEHAKATNHLILSDSAYEDLVCKKENPDMEFLKEKSAKLGHTVSNEAYSE 649

QY 392 PDKSLWRATPDMDYTKMLEEDFTYOKNSVTGFSNKGLOVKVRNKNLEIPKQHNGITIEI 451
Db 650 LEKKL--EQPSLEYL--VEHAKATNHLILSDSAYEDLVCKKENPDMEFLKEKS---AKL 701

QY 452 GDNNAFRNVDFQSKTLRKY-DLE-EIKLPLSTIRKIGAFAFQFSNNLKSFEASEDLEIKEGA 509
Db 702 GHTVVSN-EAYSELQRKYSELEKEVEQPSLAVLVERHAKATDHHHLSDSAYEDLVCKKENP 760

QY 510 FMNNRIGTLDLKDKLLIKIGDAFAFHINHIVAIVLPESVQEIGRSAFRQNGALHLMFIGNKV 569
Db 761 DVE----FLIKEKSAKLG-----HTVVSSEYSELQR----- 787

QY 570 KTIGEMAFPLSNKLESVNLS-EQKQLKTIQEVQAFSDNALSEVVL---PPNQQTREAFKR 625
Db 788 ---KYSELEKEVEQPSLAVLVEHAKATDHHHLSDSAYEDLVCKKENPDMEFLKEKS 843

QY 626 NHLKEVKGSSTLSQITFNAFDQNDGDKRFGKVVVRTHNNSHMLADGERFLIDPDKLST 685
Db 844 GH-----TVSNEAYSELEKLEQPSLAVLVEHAKATDHHHLSDSAYEDLVCKKENP 896

QY 686 MVDL-EKVLR---IIEGLDYST-RQTTQTFREMTAGKA---ILSKS---NLR 729
Db 897 VEFLKEKSAKLGHTVSNAYSELEKLEQPSLAVLVEHAKATDHHHLSDSAYEDLVCK 956

QY 730 QGEKOKFQLEAQFFLGR-VDLDKATAKEKA-----LTTKKATKNHLL----- 772

Db 957 BNPDMEFLKEKSAKLGHTVSNAYSELEKLEQPSLAVLVEHAKATNHLILSDSAYEDL 1016

QY 773 -----ERSINKAVLAVNNSAIKKANVRLEKELDLTDLVEGKGP--LAQAT 817
Db 1017 VPKCKENPDMEFLKEKSAKLGHTVSNAYSELEKLEQPSLAVLVEHAKATNHLILSDA 1076

QY 818 MVQGVYLLKTPPLPEYYIGLNVF--DKSGKLITYALDMSDTIGEQOKDAYGNPLN-V 873
Db 1077 YEELVKCKENP-----DVEFLKEKSAKLGHTVSNAYSELEK-KLEQPSLAVL 1125

QY 874 DEDNEGKHTLAVATLAD--YEGGYIKDILNSLDDKIKAIRQIPLAKYHRLGIFOAIRAA 931
Db 932 AADRILPKTPKGYLNEVNPYRKQOMEKNLKPVDYKTPFENKALPNEVKDRAK-GHN 990

QY 991 I 991
 CC :
 Db 1231 V 1231
 CC
 RESULT 3
 ALM1_SCHPO
 ID ALM1_SCHPO STANDARD; PRT; 1727 AA.
 AC 09UTK5; 013313; Q9UTT8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Abnormal long morphology protein 1 (Sp8).
 DE ALM1 OR SPAC1486.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetidae;
 OC Schizosaccharomyces.
 OC Schizosaccharomyces.
 OX NCBI_TAXID=4896;
 RN [1] SEQUENCE FROM N.A.
 RP
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Roben J., Grymonprez B.,
 RA Weltjens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer F., Moestl D., Hilbert H.,
 RA Borzym K., Langerer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaurie V., Mottier S.,
 RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.,";
 RL Nature 415:871-880 (2002).
 RN [2] SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=972;
 RX MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Godoy C.;
 RT "The alm1+ gene from Schizosaccharomyces pombe encodes a coiled-coil
 protein that associates with the medial region during mitosis.,";
 RL Mol. Gen. Genet. 262:921-930 (2000).
 RN [3] SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=20223868; PubMed=10759889;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 fission yeast cells by the use of a GFP-fusion genomic DNA library.,";
 RL Genes Cells 5:169-190 (2000).
 CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CC CO-Localizes WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC
 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 EMBL; AL133357; CAB62414.1; -
 DR EMBL; AF010473; AAB65416.1; ALT_INIT.
 DR EMBL; AB028012; BAA87316.1; -.
 DR PIR; T50073; T50073.
 DR GeneDB_SPombe; SPAC1486.04C; -.
 KW Coiled-coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1654 COILED COIL (POTENTIAL).
 FT SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 4.0%; Score 215.5; DB 1; Length 1727;
 Best Local Similarity 19.3%; Pred. No. 0.015;
 Matches 231; Conservative 198; Mismatches 422; Indels 343; Gaps 57;
 Query Match 12 ALTTVSVVTYSQEYVGLEREESVQKQEQTQASED----DWFEEDENRKTNVSKENSTVD 66
 DB 116 ALTT----ENQSLRANSELQEQSKIASQQLSIAKDQIBALQENSHLGEQVOSAH 167
 Query Match 67 ETVSD----LFSQDGNSSNSKTESWSVSSDPKQVKPKAKPEVTOEASNSNSNDASKEV 118
 DB 168 QALSDIEERKXQHMF----SSSRVKEELVQE----KSALVSDLASLQSDHSKV-- 215
 Query Match 119 PKQDTASKETLETSTWEAKDFVTRGDTLVGFSKSG--INKLQTSQTSHLVLPSHAADG-- 173
 DB 216 ----CEKLEVVSSRQVQDLEK---LAGLAQONTTELNEKIQLEQEKRSNYSQSDGNIS 264
 Query Match 174 ----TQITQVASEFAFTPDKTTAIAEYTSRIGENGKPSRLDID---QKEIIDEGEIF 222
 DB 223 NAVQLTK----LTIPIVNGYKSIGQ--DAFVDNKNIAEVN----LPE 257
 DB 265 KILETDPTSIKELEEEVETQKRITALWE----SKSSELQSEVAALQEKLTQSQSLY 316
 Query Match 317 N--NVTEELNNNKQQLLISNRLQELQEKYDVSSELQVKENKNTSVAGVGLFSPLAQ 374
 Query Match 258 SLETIS--DYAFAHMSLKVQLPDNL----KVICELAFFDNQIGGKLVLPRHLKL 307
 DB 375 KLSAVONPEFFSFTKVYSDNMKLLQOKVSSLKLQLDRILTKNFESFCEQV--KQRIK--WVKQ 430
 Query Match 308 AERAFKSNRIQTVEFLGSKLKVIGEASFDQDNILRNVMPLQGLEKIBESAAFTGNGPGEDEHYN 367
 DB 431 QRSEIIVRNNTY-MNFLSESLET----SNNNLTKVQAEELLSTKMRQEA----YL 475
 Query Match 368 NQVVLRTTRGQNPHQ--LATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNQSVTG-- 422
 DB 476 QLTASRTQCSDSLRSREVICLMAELDHUNETKS--RNVPATQVA-LDE--YAQNPTASE 529
 Query Match 423 -FSNKGLQ----KVRRNKLEIPKQHNGITITEIGDNAFRNVDQSKRLKYDDEEIKL 476
 DB 530 TLVNLKELANFESSTIKEAVSKTLELRKVALEC----DVEIQKQTQYQQISNAVKE 580
 Query Match 477 PSTIRKIGAFQSNLKSFEASEDLEEIKEGAFMNRR----IG 516
 DB 581 NSNT----LSEQIKNLSELSNNSKIKNESSLNERNLKEMLATSSRLSILSHNSAG 632
 Query Match 517 TLDLKDQLIKIGDAAFHINHIYATVLP--SVOEIGRSAFRQNGALHLMFIGNKVKTIG 574
 DB 633 NID--DKMKSIDESTRELEKNYEVNEMTAQE--SLSKRNQDL----LSE 676
 Query Match 575 MAFLSNKLESVNLSSEQQLKT----IEVQAFSDNALSEVVLPPNQI--REEAFKR 625
 DB 677 MEAIRKELE--NSKVQQQLSTDRLTNANNDVEAFKKEAKELRSINQNLQDIISRQDQRAS 734

QY 626 NHILKEVKGSSTLSQITNAFDQNDGDKRFCKVVRHNNHMLADGERFTIDEDPKLST 685
 RA :|:|:|:|:|:|:|:
 Db 735 KFAEELIIVNLSLAERLKGELNSKGEKDLRKRT-----QERLISENDKL-- 778
 RA :|:|:|:|:|:|:
 QY 686 MVDLEKVLKLTIEGLDYSTLROTTOTOFREMTAGKALLSKSNLROGEKQKFLQEAQFFLG 745
 RA :|:|:|:|:|:|:
 Db 779 LAERERLMSLVSDL-----QTFLNO-QQISDAAR-----KVKFSEKE--- 815
 RA :|:|:|:|:|:
 QY 746 RVDLDKATAKAEKALVTKATKNGHLLERSJNKAVLAYNN-----SAIKKANVRL 796
 RA :|:|:|:|:|:
 Db 816 --SLSLSLQKLES--NEKMSNDLHSIQLSKSLEKSGIEYSSRIKTLIMLEKOSLSEDNRKL 871
 RA :|:|:|:|:|:
 QY 797 E-----KELDLITDIVEKGKPLAQATMVO-----GVTLLKTPL-PLPEYVIGL 838
 RA :|:|:|:|:|:
 Db 872 DNOQOMMETIKLQELNGVIELEKORFSTLEAKFTQOKNTSYSEREALLESSLSDQSKHTSL 931
 RA :|:|:|:|:
 QY 839 NVYFDKSKLIVYALDMSDTIGEGQ---KDAYGNPLINVDNEGYHTLAVTLADYEGI 894
 RA :|:|:|:|:
 Db 932 ESQVNYSILRNTEQLOAQASKLAEEMVERVKTExDEYRLOTSESLEKH-LKITSQRIVI 990
 RA :|:|:|:|:
 QY 895 YIKDILNSSLDKIKAIROQIPLAKYHRLGIFQAIRAAEAD---RLLPKTPKGYLNEVPN 951
 RA :|:|:|:|:
 Db 991 LQDEIASSL-----RCENTIKDSETRVAILLEENKHNLLESS 1029
 RA :|:|:|:|:
 QY 952 YRKQMEKNLKPVYKTP-----FNKALPN 977
 RA :|:|:|:|:
 Db 1030 HRNAEKQHLEKENDYKQQLLIVTEDLRKTRDYEKELLRHADARSTIQLKREDYTKALEQ 1089
 RA :|:|:|:|:
 QY 978 -EKVDGDRRAK-GHNTNAETNNSVAVTPIRSBQOLHKSQSDVNLFQTSKKN-NF 1028
 RA :|:|:|:|:
 Db 1090 VEDLNKEIALKAGINESQPFPISEKEDPLRQEVTVLKQNAMLLTQLOSSNLF 1143
 RA :|:|:|:|:
 RESULT 4
 PCP1_SCHPO ID PCP1_SCHPO STANDARD: PRT; 1208 AA.
 AC 092351;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Spindle pole body protein pcpl.
 PCP1 OR SPAC6G9.06C.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21848401; PubMed=11859360;
 RA Flory M.R., Morpew M., Joseph J.D., Means A.R., Davis T.N.;
 RT "Pcp1, a Spc110p-related calmodulin target at the centrosome of the
 fission yeast Schizosaccharomyces pombe.";
 RL Cell Growth Differ. 13:47-58 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11864908;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Squires J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jageis K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squires R., Squires S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonpre B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

Ra Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 Ra Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 Ra Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 Ra Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 Ra Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 Ra Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 Ra Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Ra Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 Ra Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 Ra Shpakovski G.V., Usery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe.",
 RL *Nature* 415:871-880 (2002).
 CC -!- FUNCTION: Spindle pole body component that binds calmodulin.
 CC Overexpression of pcpl causes the formation of supernumerary SPB-like structures and disrupts both mitotic spindle assembly and
 CC chromosome segregation.
 CC -!- SUBCELLULAR LOCATION: Spindle pole body.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR GeneDB_SPombe; SPAC6G9.06c; -
 DR Calmodulin-binding; Coiled coil.
 FT DOMAIN 151 375 COILED COIL (POTENTIAL).
 FT DOMAIN 387 803 COILED COIL (POTENTIAL).
 FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
 FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;
 Query Match 3.9%; Score 209; DB 1; Length 1208;
 Best Local Similarity 19.4%; Pred. No. 0.018;
 Matches 220; Conservative 170; Mismatches 367; Indels 378; Gaps 52;
 QY 20 TYSQEVYGLEREESVQKQETQSA----SEDDWFEEEDNERKTNVSKENSTVDETSDLFS 74
 RA :|:|:|:|:|:|:|:
 Db 93 SFDPNSYGL---SAISKQATOALSIQGNDSY---DVSKLTDLSK-NSBIDHTDGEI-- 143
 RA :|:|:|:|:|:|:
 QY 75 DGNNNNNSSKTESVVD-----PKOVPKA---KPEVTVQEASN 108
 RA :|:|:|:|:
 Db 144 PANALTREQEVKLYSRENFGRLIKTVCLERKLESMAPEQIKEAVKDNDVLAERAN 203
 RA :|:|:|:|:
 QY 109 SSNDASKVVEPKQDTASKETLETSTWEAKDFVTRGDTLIVGFSKSGINKLSQL-----T 161
 RA :|:|:|:|:
 Db 204 LQLQKRTESLQLQSKEDQNKFLK-----EKVYLSK-----VNDVEQSONVKVFT 248
 RA :|:|:|:|:
 QY 162 SHIVLPSHAADGTQLTQVA-SFAFTPDKKTALEY-----TSRLGENGKPSRLIDQK 213
 RA :|:|:|:|:
 Db 249 ERIFLENALEKVKOREKDKSLSTMEDKSNKEVDYVEIRQLQNLDELSE--ELDVAQ- 305
 RA :|:|:|:|:
 QY 214 EIDDEGETFNAYQLTQLTIPNGYKSIQDAFVNKNIAEWNLPE----- 257
 RA :|:|:|:|:
 Db 306 DLITTEKE---DEIATLKRQTEKEENSSAFENEENSSYVHQLQDYLQAKCDEFADRI 361
 RA :|:|:|:|:
 QY 258 -----SLETISDYAHMSLKVQKLDPNLKVIGELAFFDNQIGGKLYPRHLIKLAAERAF 312
 RA :|:|:|:|:
 Db 362 QVLTDADKEKENQIMHESEASIGLTDSMQV-----H-TLQEQLH 400
 RA :|:|:|:|:
 QY 313 KSNRIQTVEFLGSKLKVIGE---ASFQDNMLRNMLPQDGLKIESEAFGTGNPGDEHNNQV 370
 RA :|:|:|:|:
 Db 401 KAN-BEIEFLHDOISRMNEEGKNFEDIMLQFRSLEERDVLQESKQ-----LEDDNNSL 454
 RA :|:|:|:|:
 QY 371 VLRTT-GONPHOLATENTYVNPDKSLWRATPMDYTKWLEDFTYQKNSVTGFSNKGLQ 429
 RA :|:|:|:|:
 Db 455 RLMTSSIGNOTESLRTRNREIDEKHLRLLASKNSDKALAE-----TNIRLQ 502
 RA :|:|:|:|:

QY 430 KVRRNKNLEIPKQHNGITTEIGDNAFRN-----VDFQSKTIRKYDLE 472
 503 EV--TKELETLRMKNNSNDLNEIHDLRREENEGTLKIDSITKEKDLINELQRQIKSYEVN 560

QY 473 EIKLPSTIRK-----IGAFAFQSNLKSFEASEDLEEIKRGAFFMNRIGHT 517
 561 VSELNGTIDEYRNKLUKQKEETYNEVMNAFOYKNDLRRFRESINKLQDREKEITSN--- 616

QY 518 LDLKDKLIKIGDAAFHINHIYAIVLPESYEIGRSAFRONGALHLMFIGNWKVKTIGEMA 577
 617 LEKKNLV-----SSIRETVALEKERESIKYKLSGN----- 648

QY 578 LSNKLESVNVLSE-----QKQLKTI--EVQAFSDNALSEVVLPPNQLTREEAFKRN 626
 649 -AKDLDNTNLMEILNDKISVLRQLTDVKOELDVSEEEEREAIVAGQKLSASFEILMSNEK 707

QY 627 HLKEVKGGSSTLSQITFNAFDQNDGDKR-----FGKKVV---VRTHNNSHMLADGERFI 676
 708 QALELKYSLLKNEL--INA--QNLLDRREELSELSKLFEERKIRSGSNDDEKNEINV 764

Db 765 LNSELADKLAQIRHLESKMEQLDKLVLHHLNRGTEEANIEENAVKRLCLIMGCDYSSVI 824

QY 677 ID---PDKLS-----STMVDLEK-----LKIIEGDDYSLR- 705
 706 -----OTTOQFREM-----TTAGKALLSK 725

Db 825 LQIVSQIEHFVNQOQTIRSLKQELRHDVQFSGKKEQBLRSPEKFGLIGTETKHDILAQ 884

QY 726 SNLROGEKQKFLQE-AQFFLGRVD-----LDKAIAKAELAKALVTKKATK 767

Db 885 RNRNVSEKMDLLENAQKFSSPDRKNGVLYPSEHTSKIEYLEKTIEDLKLAI--QDELK 942

QY 768 NGHLLERSINKAVLAVNNSAIK-KANVKRLEKELDLTD-----LVEGKG 811

Db 943 NRNLIMMDIS-----SYNKQTTKLOEKIKWLERERSILIDELEYRSNQPNYQMLNVQDKN 998

QY 812 PLAQATMVGVYLLKTPPLPEYYIGLNVY---FDKSGKLIYA--LDMSDTIGEQKDAY 866

Db 999 ELEER-----LKEIQLKELEVNNHFMKQAEIMTSNVTDESQMLMLTREAL 1044

Qy 867 GNPILINVEDNEGYHTLAVATLADYEGLYIKDILINSSLDKIKAIROQIPLAKYHRL 921

Db 1045 QSKINNIDH---LSTIILERNRKEYKSL-----LDDYNQLR---ARYKNL 1082

RESULT 5

MLP1_YEAST STANDARD; PRT; 1875 AA.

ID _MLP1_YEAST
 AC Q02455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MLP1.
 GN MLP1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=93247549; PubMed=8483450;
 RX Koellinger R., Nguyen T., Chen E.Y., Botstein D.;
 RT "A new yeast gene with a myosin-like repeat structure.";
 RL Mol. Gen. Genet. 237:359-369(1993).

RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.I.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three new open reading frames.";

RQ yeast 9:1349-1354(1993).
 CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA repair.
 CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -!- CAUTION: Ref. 2 misquotes the gene name as "MLP1".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 Db EMBL; L01992; AAA34783.1; -.
 DR EMBL; X73541; CAA51948.1; -.
 DR EMBL; 228320; CAAB2174.1; -.
 DR GO; S38173; S38173.
 DR GermOnline; 140074; -.
 DR SGD; SGD001803; MLP1.
 DR GO; GO:0005635; C:nuclear membrane; IDA.
 DR PIR; S38173; S38173.
 DR GO; GO:0005654; C:nucleoplasm; IDA.
 DR Coiled coil; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1)
 SEQUENCE 1875 AA; 218455 MW; 683A0D34C9056867 CRC64;

Query Match 3.9%; Score 209; DB 1; Length 1875;
 Best Local Similarity 19.2%; Pred. No. 0.033;
 Matches 221; Conservative 194; Mismatches 375; Indels 362; Gaps 55;

QY 32 ESKQEQTQASSE-----DDWFEEDNERKTKNSKENSTVDETVSDL---FSDGNSN 79
 107 ESENENKMKLSELEFVKRKLDLITEEKKETQSNQQRTLKILDERLKEIELVRVNNRNS 166

QY 80 NSSKTESVVSDFPKQVPKAKPEVTOEASNSSNDA-SKVEVPKQDTASKKETL--ETSTWE 136
 Db 167 SECKKLKSTIMDLE-----TKQOQYITNDLNSRTELERK--TQELTLLQSNNDWL 214

QY 137 AKDFVTRGDTLVGFS-----KSGINKL-----SQTSHIVLPSHAADGTQITQV- 179

Db 215 EKELRSKNEQYLSYRQTDKVIDLIRNLRNLDQFQMRERTNDVLIKQKNNELSKSLOEK 274

QY 180 -----ASFAFTPDKK-----TAIAEYTSRIGENGKPSRDLIDQK 213

Db 275 LLEIKGLSDLSLNSEKQFSAEMLKQRLVLILLESQNLNAKEELNSIRENTAKVIADD SK 334

QY 214 EIIDEGE-IEFAYQLTQTKLITIPNGYK-----SIGQDAFVDNKNIAEWNLPESLETISDYA 266

Db 335 KQTPENEDLJELQLTKEKLAQCEKECLRLSSITDEADEDNENL-----SAKSSSDF- 386

QY 267 FAHMSLKVQLPDNLKVIGELAFFFQIGGKLYLPRHLIKLAERAKF---SNRIQT-VEF 322

Db 387 -----IFLKKQLIK--ERRTEKHLQNIETFIVE 413

QY 323 LGSKLKVIGEASFQDNNLNRNMLPDGLEKIESAFTGNGPDEHYNQ--VVLRTRTGQNP 380

Db 414 LEHVKVPIIN--SFKERT-----DMLENELNNAALLHTSNEKNAVKELNAHQ 461

QY 381 HQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSTGFSNKGQK--VVRNKNLE 438

Db 462 KLVCENDLQTLTKO----RLDCRQIQ-YLLTINSVNSNDSKGPLRKERBQFIQNI- 512

QY 439 IPKQHNGITTEIGDNAF--RNDVDFQSKTIRKYDLEEIKLPSTIRKIGAFAFQSNLKS 495

Db 513 --MQEDDSTITESDSQKVTERLVEF--KNIIOLOQEKNAELLKVRNL-ADKLESKEKKS 567

QY 496 FEA---SEDEEIKEGAFMNRIGTLDLKLKIGDAAFHINHIYAIVLPEVQEI 550

Db 568 KQSLQKIESETVNEAKE-AITLKEKMDLESRIE-----LQKEEL- 610

QY	840 VYFDKSGKLIYALDMSDTRI----GEGOKDAYGNPILNVDE 875
Db	909 LIDDKTLSEKVRLOTIDKLLQEQTQKQAEATENPSVKTED 948
QY	RESULT 7
SCA4_RICAK	SCA4_RICAK STANDARD; PRT; 998 AA.
ID	SCA4_RICAK
AC	Q9ATX9;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Antigenic heat-stable 120 kDa protein (ps120) (Protein PS 120) (Fragment).
DE	(Protein PS 120) (Fragment).
GN	SCA4 OR D.
OS	Rickettsia akari.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia.
OX	NCBI_TaxID=786;
RN	[1] SEQUENCE FROM N.A.
RA	Sekeyova Z.; Roux V.; Raoult D.;
RT	"Phylogenetic analysis of Rickettsia spp. by comparing sequence of the gene D' coding for an intracytoplasmic protein.";
RL	Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC	-1- SUBCELLULAR LOCATION: Cyttoplasmic (Probable).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	or send an email to license@isb-sib.ch).
DR	EMBL; AF213016; AAK30691.1; -.
KW	Antigen.
FT	NON_TER 1 1
FT	NON_TER 998 998
SO	SEQUENCE 998 AA; 109328 MW; FCEBA3AC62DE5BDS CRC64;
Query Match	3.8%; Score 203; DB 1; Length 998;
Best Local Similarity	18.9%; Pred. No. 0.027;
Matches	220; Conservative 176; Mismatches 405; Indels 360; Gaps 55;
QY	6 LKTLALALTTVSVWTYSQEYVGLEREESVKQEQTOSASE-----DDWFEEDNERKT 56
Db	57 ISTLSGTISTDDQI--SDPITKAVREIIQQQKDEIAEQILKDLAALVDRDLAEQKRKEI 114
QY	57 NVSKENSTVDETVDLPSDGMSNNSSKTESVVSDFPKQVPKAKPE-----VT 103
Db	115 EEEKEK--DKKLSVFF--GNPAN----REFIDNALEKPELKKLESIEITGYKNILLT 164
QY	104 QEASN-----SSNDASKVEVKQDFTASKKETLETSTWEAKDFVTRGDTLV 148
Db	165 YSAANGYHGGFKPVQWENQISASDL-RATVVKNDAGDELCTLNFTVKTPFTVA---- 218
QY	149 GFSKSG-----INKLSQTSHLVLP SHAADGTQLTQVASFAFTPDKKTAI 192
Db	219 --KKDGTQVQINSYRAIDFPPIKLLKADGSMHLSMVALKADGTVK-----PSKDRAV 266
QY	193 ---AEYTSRRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTI PINGYKSIGQDAFVDNKN 249
Db	267 YFTAHYEE--GPNGKP----- 280
QY	250 IAEVNLPELETISDYAFAHMSLKVQLPDLNKVIGE---LAFFDNQIGGKLY---LPR 302
Db	281 -----QLEKEISSPQPLKFAGDGPDAVAYIEH--GGEIYTLAVTR 317
QY	303 HLIKLAERAFFSNRIQTV--EFFLGSKL-KVIGEASFQDNRLRVMPLDGLER--IESEA 356
Db	318 GKYKEMMKEVELHOGQSVDSLQITAEDILTKVQGRS--QETIQLPITPNQELKSSIEPT 374
QY	357 FTGNPGDHEHYNNOVVLRTTGTQONPHQLATENTYVNPDKSILWRAPEM-----DTKMLEE 411
Db	375 TTQVPPITPASQPV--HTETSQMP----QSQQVNP--NLFNAATALCSMSQDLNLNVNA 425
QY	412 DFTYQK--NSVTGFSNKGLQKVRRNKNLEIPKQHNGITTE--IGDNAF----RNVD 461
Db	426 GLTKEKDGTQIDLINEATAILNN----EKEQANFILTKNMNNALTPDTKVARVNA 482
QY	462 QSKTLR-KYDLEEIKLPSTIRKIGAFAFQFSNNLKSFEASEDLEEIKE----- 507
Db	483 VLETIKNQNQDTPDIEKSKMLA TVALTNSENTLPKQKQOMLEKAVDVDSLFSKDDTSRAV 542
QY	508 -----GAFMNNRIGTLDLKDMLIKIGDAFAFHINHIVIAIVLPE SVOEIGRSAFRONGALH 561
Db	543 AIDGITGAVIKSNLJSTKDKGTTMLIAVGDKV-----NASELSNAEKQQ---- 584
QY	562 LMFIGNKVKTIGEMAFLSNKLESVNLSEQKQLKTIEVQAFSDNALSEVVLPPNLTIREE 621
Db	585 --LIGSVLKKGVETKLS-----PEQQQL-----MQQNLDDKITA 617
QY	622 AFKRNHILKEVKGSSTLSQITFNAFDQNDGDKRFGKKVVRTHNNSHMLAD--GERFIIDP 679
Db	618 QTKNNDNITEVQG--TIANPAFTIATAIQKVTTKL----DSPITAEIKE----- 664
QY	680 DKLSSSTMVLDLEKVLKITEGLDYSTLRLQTQFREMTTAGKALLSKSNLRLQGEKQKFLOE 73.9
Db	665 -----TLESTIKIVA--ESPLNVQDKTDI--VKGMGEATASHRTMAPTKKIAAE- 710
QY	740 AQFFLGRVLDKAIAKAEKALVTKKATKNGHLLERSINKAVLAYNNSAIKKANVKRLEKE 799
Db	711 -----SVEITGVAKSITDLEDKLMWTKGLVTDGIVYEDKA-----NPETTSEMMKAVSKG 757
QY	800 LDLLTDLIVEKGKGPLAQ-----TMVQGV-YLIKTPPLPEVYIGLNVYFDKSG 846
Db	758 VDNSTAIPEDKOALKDAASEAALDRATONFTEGLKGQNLDEPKPRDDIY-----NKAQ 810
QY	847 KLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHTLAVTLADYEGLYIKDILNSSLKD 906
Db	811 DIAVALK-----NVVTTVLDANPEKREVSEEEVMKTSILNDISKIAEK 856
QY	907 IKAIRQIPLAKYHRLGIFQAIRNAAEADRLLPKTPKGYLNEVNPYRKQ-MEKNLKPWD 965
Db	857 VNNLRLRAM-LSPDSSL----KTLLEBKKAEATKKVDELVKEFFGTSSTEEQQSFIQANL-ID 910
QY	966 YKT-----PIFNKALPNEKVDGDRRAKGHNINAE---TNNSVAVTPI-RSEQQLHKS 101: 911 DKLTSKEVRLQDIDKLLQEQAQKRAEAIKNPVNKTEDLRVSGQSAALKPISNDEPDIKT 970
Db	1014 QSDVNLQPSSKNFIVYELG 1034
Db	971 KMVVGDRVNKDN-----KIMG 989
QY	RESULT 8
RPB1_PLA8B	RPB1_PLA8B STANDARD; PRT; 2869 AA.
ID	RPB1_PLA8B
AC	Q00798;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-APR-1993 (Rel. 25, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Reticulocyte binding protein 1 precursor.
GN	RPB1.
OS	Plasmodium vivax (strain Belém).
OC	Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=31273;
RN	[1] SEQUENCE FROM N.A.
RP	MEDLINE=92315338; PubMed=1617731;
RR	Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT	"A reticulocyte-binding protein complex of Plasmodium vivax merozoites.";
RL	Cell 69:1213-1226(1992);
CC	-1- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to

CC -!- human reticulocyte cells.
 CC -!- SUBUNIT: Homodimer (Potential).
 CC -!- SUBCELLULAR LOCATION: Membrane-bound.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M88097; AAA29743.1; -.
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 18 2869
 FT DOMAIN 18 2807
 FT TRANSMEM 2808 2826
 FT DOMAIN 2826 2869
 FT SITE 1030 1032
 FT SITE 2599 2601
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE42205EBCFF CRC64;

Query Match 3.8%; score 202.5; DB 1; Length 2869;
 Best Local Similarity 17.9%; Pred. No. 0.12; 406; Indels 385; Gaps 50;
 Matches 211; Conservative 178; Mismatches 886

QY 4 KHLKLTALLATT-VSVVITYSQEVYGLEREESVKQEQTQSASEDDWF----ED---- 51
 Db 827 KKMKEFSDAFSTKFEALQNSM0QYNOEQGDATEKHKQNRSEKEEEYFKNESVEEDLSREET 886
 QY 52 - - - - - NERKTKNSKENSTVDETVDSDLFSDGNSSNNSKTESVVSDFPKQVPA- 98
 Db 887 EEQEYTKHHKNFSSRRKGELISAETINMREVINKTESQNLNYGVIEKFSLIGDQNEVSTAK 946
 QY 99 - - - - - KPEVTQEAQNSNSNDAKVEVKQDGTASKETLET--STWEAKDFVTR----G 144
 Db 947 ALKEKIVSDSILRKDIDQYETEF-KEKTSAVENTVSTIQSLKAIDSILKRINGSINNCKY 1005
 QY 145 DTLVGFSSKGNGNKLSTQTSHLVLPSHAADGTLQTOVASFAFTPDKTATAEYTSRLGENGK 204
 Db 1006 NTDDIDLRSKIKTIREEVQKEMPKRGDKGCENTTALLKSLRKDGKINEKLN----DGR 1061
 QY 205 PSRIDIDQKETIDEDEGIFNAYQLTQKLTIPNGYKSIGQDAF--VD----NKNIAEVNLPE 257
 Db 1062 LNSLDTKEKDLK---FYSSESKSKIHLKDQKG-PQDPLNRIDEWEDIKRDVDELNV- 1114
 QY 258 SLETISDYFAHMSIKQVKLDPDNLKVGIGELAFFDNOIGGGKLYLPRHLIKLAAERAFKSNRI 317
 Db 1115 - - - - - NYQVISE----NKV----TLPKNSV 1132
 QY 318 QTVEFLGSKLKVIGEASFQDNNLRNVMLPDGLEKIESEAFTGNPGCDEHYNNQOVLRTRTG 377
 Db 1133 TYTEAMHSINTV----AHGITSN----KNEILKSVKEV 1163
 QY 378 QNPHQLATENTYVNPDKSLWRATPDMDYTKMLEEDEFTYQKNSVITGFSNKGLOQVRRNKL 437
 Db 1164 EDKLNLVEN----EDYKVKNPE--NEKOLEAIRGMS- 1196
 QY 438 EIPKQHNGITITEIGDNAFRNVDQFOSKTLRKVDLEEIKUPSTIRKIGAFAFQSNNLKSF 497
 Db 1197 - - - - - KLKEV----INKHVSMTOLESTANTL----KSNAKGKE 1227
 QY 498 ASEDLERIKEGAFMNMRIGTL-DLKDKLIKIGDAAFHINHIYAIYVLPESVQELGRSAFQ 556
 Db 1228 NEHDLEB----LNKTKGOMRDIYEKLKRCAE----LKEGTVNELKDAEK 1270
 QY 557 NGALHLMFIGNKVKTIGEMAFLSNKLESVNLSEQKQKLTIE----VOAFSDN 604
 Db 1271 - - - - - ANKVEPEPERNITIGHVLERITVEKDKAGKVVEEMMSLKTKEKLIQFETSDD 1321
 QY 605 ALSEWV--LPPNQQTIR--EEAFKRN----HLKE-VKGSSTLSQITFNAFDQNDGDKR 653

Db 1322 SQNELVTTSTIKHLENAKGYEDVIRNEEDSIQIREKAKSLETDEM----- 1368
 QY 654 FGKVVVRTHNNSHMLADGERFIIDDDKLSSSTMVDEKVKIITEGLDYSTL----RQ 707
 Db 1369 - - - - - KKLVQQVMNLQSAIQGNAGI----SKELNELKGVIELLISTNNSILEYVKKNS 1420
 QY 708 TQTOFREMT---TAGKALLSKSNLROGEKQKFLQEAQFLGVRDIDKATAKAECALKVTK 763
 Db 1421 BSVRFSQLANGEPTKAEGEEKNASARLAEAKLKEQIVKOLDYSDIDDKVKKIE---- 1474
 QY 764 KATKNGHLLERSINKAVLAYNNSAI----KANVKRLEKELDLTDLVE 808
 Db 1475 - - - - - GIKREILKMKESALTFWEESEKFKQMCSSHHMENAKEKKIEYLKN--N 1521
 QY 809 GKGPLAQATMWQ---GVYLLKTPPLPEVYIGLNVYEDKS----GKLTYALDMSDTIG 859
 Db 1522 GDGGKANITDSDOMEEVGNVYNSKA---EHAFHTVEAQVQDVKTKAFCESIVAYVTMDNLFN 1577
 QY 860 EGQKDAYGNPILNVEDNEGYHTLAVATLADYEG----- 893
 Db 1578 ESLMKEVKVKCERKNDAEAKYS---AKLKPYDGRKARVSENERKISELKEKAKVKEKE 1633
 QY 894 -LYKDILNSLDKIKAIRO-----IPLAKYHRLGIFOAIRNAEAADRLLPKTPKG 944
 Db 945 YLINEVPNRY--KKOMEKNLKPVDYKTPIFNKPALPNEKVGDRAAKGHNIAE---TNNS 998
 QY 999 VAVTPIRSEQQLHKSQSDVNLPO---TSSKNNFIVYILG 1034
 Db 1691 AEKSDLKVKAAKESYENOLETQ----NEM----SRINVEEGSLTDID 1730
 Db 1731 KKTDIENDLILKMKKQYEEGLQKIKENADKRRKSNFELVG 1770

RESULT 9
 SCA4_RICPR ID SCA4_RICPR STANDARD; PRT; 1022 AA.
 AC Q9ZD49; Q9AR36; Q9ZD48;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
 DE (Protein PS 120).
 GN SCA4 OR RP498/RP499.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alpha-proteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid B;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.B., Zomodropour A., Andersson J.O., Podowski R.M., Naeslund A.K.,
 RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RA "The genome sequence of Rickettsia prowazekii and the origin of
 Rickettsiaceae; Rickettsiae; Rickettsia.
 RL Nature 396:133-140 (1998).
 RL [2]
 RP SEQUENCE OF 11-1016 FROM N.A.
 RA Sekeyova Z., Roux V., Raoult D.;
 RT "phylogenetic analysis of Rickettsia spp. by comparing sequence of the
 RT 'gene D' coding for an intracytoplasmic protein";
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- CAUTION: REF. 1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 234.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AJ235272; CAA14951.1; ALT_FRAME.
 DR EMBL; AJ235272; CAA14950.1; ALT_FRAME.
 DR EMBL; AF200340; AAK31305.1; -.

KW Antigen; Complete proteome.

FT CONFLICT 11 15 EFDPL -> RPLGLV (IN REF. 2).
 FT CONFLICT 365 365 H -> Y (IN REF. 2).
 FT CONFLICT 413 413 MISSING (IN REF. 2).
 FT CONFLICT 957 957 G -> R (IN REF. 2).
 SQ SEQUENCE 1022 AA; 114410 MW; 03230E3A663A9622 CRC64;

Query Match 3.8%; Score 202; DB 1; Length 1022;
 Best Local Similarity 18.8%; Pred. No. 0.031;
 Matches 207; Conservative 196; Mismatches 400; Indels 298; Gaps 54;
 Matches 207; Conservative 196; Mismatches 400; Indels 298; Gaps 54;

QY 28 LEREESVKQE---QTQSASEDDWFEEDNERKTNVSKENSTVDETVSDLFDSDGNSSNNSSK 84
 Db 85 LEKQRDILREYFVNTPELAEQIAKEEDDRKFRAFLSNQDNYALINKAFEDTKKKLEK 144

QY 85 TE-----SVVSDPKQVPKAKPEVTOEASNSNDASKVEVKQDTASKKETLETSTWE 136
 Db 145 AEIVGYKVNVLSTYSVANGYQGGFQPVQWENQVSASDL-RSTVVKNDEGEELCTLTVK 203

QY 137 AKDFVFT--RGDTLVGFSK-----SGINKLSQLTSHLVLSHAADGTQLTQVASFAFTP 187
 Db 204 TKDLIVAKQDGTVQVQINSYREINFPPIKLDKANGSMHLSVALKADGTAKDKAVYFT-- 261

QY 188 KTKTAAEYTSRLGENGKPSRLDIDQKEITIDEGEIFNAYQLTKLTIIPNGYKSIGQDAFVDN 247
 Db 262 -----AHYEE--GPNGKP----- 272

QY 248 KNLAEVNLPESETISDYAFAHMSLKVQLPDLNLKVIG----ELAFFDNQIGGKLY---L 300
 Db 273 ------QLKEISSPQPLKFVGTGDDAVAYIEH--GGEIYTAV 307

QY 301 PRHLIKLAERAFAFKSNRIQTV---EFLGSKLKVIGEASFODNNLRNVMLPDGLEKIESEAF 357
 Db 308 TRGKYKEMMKEVALNHGQSVALSQTAEDLTHVQGSPSHETH--KPIIIPN--QELESSI- 362

QY 358 TGNPGDEHVNQV-----VLRTRTGQNPQHOLATE-----NTYVNPDKSLWRATP 401
 Db 363 -----EQHTSQQQVPPITTFNKLQLQPKISQ-IHQLQPOQAQSSGIPNPVLNAANALSTSMQ 416

QY 402 DM--DVTWKLEEDFTYQKNSVTGFSNKGLQKVRRNKLEI-PKQHNGITITEIGDNAFRN 458
 Db 417 DLLNNINSYLTKNQDINKQS--DLIKEAAIAILNNKKSDFAEKQY--IIDLAKNIFSN 471

QY 459 VDFQSKTIRKYDLEEIKLPSTIRKIGAFAFQSNNLKSFEASEDLEIKEGAFMNNRIG-- 516
 Db 472 KDIIDA-----KVNVNNTL-----LETIONDONTLDIJKSKILEDVAIT 512

QY 517 ----TLDLKDK---LIKIGDAAFHI-NHIYIAIVLPESVOEIGRSAFRONGALHLMFIGNK 568
 Db 513 LNSENIELKOKQQILEKVVVDIGLISIKDDISRVAVDSIMD---TVIKSN-----TANE 562

QY 569 VKTIGEMAFLS-----NKLESVNLSEQKQL----KTIEVOAFSDNALSEVVLPPNLQTI 618
 Db 563 DK---EKIFITVFDQINSYEFNSVAKQKLLDSILKKTAETQVLSPE--QQQLMNLQNLDNI 617

QY 619 REEAFKRNHLKEVKG-----SSTLSQITFNAFDQNDGDKRFGKKVVVRTHNNSHMLADG 672
 Db 618 TTEHTKRTDIEKVNNTILLEPLSNTALKTT-----NIQVMTSN----- 654

QY 673 ERFITIDPDKLSSTMVDELEKVLKIEGLDYSTLROTTQFREMTAGKALLSKSNLROGE 732
 Db 655 ---VLD-----SPVQIEMKSKLQVVT-KTVAESALVEPKDKTEIVKG-IGKTIIVTHSD 703

QY 733 KQKFELQEAQFFLGRVVDLKAIAKAECALVTKKATKNG--HLLERSINKAVLAVNNSAIK 789
 Db 704 TSLPLHDKVWIMGSV--AKGIVESKNDLDRLLIAGLVDGIVYEAKGDNAWHAISSMIA 761

```

QY 790 KANVKRLEKE-LDLTDLIVEGKGPLAQATMVGVYLIKTPPLPEYYIGLNVYFDKGKL 848
Db 762 NSNINQSEKEALKRSQDVSEK-----VLDKEIQ-----NLDRELKAQN 800
QY 849 TYALDMDSDTIGEGQKD--AYGNPILNVDDEDNEGHTLAVATLADVEGLYIKDILNSSL 905
Db 801 INESKLHDDIYNKTQDVANALKNVITVLDNSGQRGVSSEAPKKVSSL-LNDISKRTIE 859
QY 906 KIKAIIRQIPLAKYHRLGIFQAIRNAAEADRLLPKPGYLN----EVPNYRKQOMEKN 960
Db 860 KINNLRAM-LSQDGNLKTFEKKD--EATKKVDELVKAFDNKSSTEEQQNFIKSNL-- 912
QY 961 LKPVVDYKTPIFNKALPNEKVDG----DRAAKGNINAETNNSVAVT----PI-RS 1006
Db 913 ---IDNKT--LSREIRLQIIDNLKAQKRAETIENLSAKTEDVRVISGKSELKPIQD 967
QY 1007 EQQLHKSQSVDVNLQQTSSKNN 1027
Db 968 EPVIQAKMVVERDRVDIKDN 988

RESULT 10

TE95_TETTH
ID TE95_TETTH      STANDARD;      PRT;      872 AA.
AC Q94819;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Telomerase component p95 (EC 2.7.7.-).
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95292335; PubMed=7774009;
RA Collins K.; Kobayashi R.; Greider C.W. ;
RT "Purification of Tetrahymena telomerase and cloning of genes encoding
the two protein components of the enzyme.";
RL Cell 81:677-686(1995).
CC -!- FUNCTION: Ribonucleoprotein DNA polymerase that catalyzes the de
novo synthesis of telomeric simple sequence repeats. P95 contains
some or all of the template-independent primer DNA-binding site
CC -!- SUBUNIT: Telomerase consist of two subunit, p80 and p95 that form
a 1:1:1 complex with the 159 nt telomerase RNA.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U25642; AAC46602.1; -.
DR PIR; S55940; S55940.
KW Transferase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
KW DNA-binding.
SQ SEQUENCE 872 AA; 102946 MW; 78AFF4CD829651FA CRC64;

Query Match      3.7%; Score 200; DB 1; Length 872;
Best Local Similarity 20.3%; Pred. No. 0.031;
Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;
QY 14 TVSVWVTVSQEEVYGLE-----REESVK-----QEQTQSASEDDWFEEDNERK- 55
QY 17 TNLDFVQLQNLLEVYKSOIEHYKTQQQIKEEDLKLKFKNQDQDGNGNDDDEENNSNQ 76
QY 56 -----TNVSKENSTVDETVDLFDGNSNNSSKTESVSDPKQVPKAKPEVTEASN 108

```

QY 77 QELIRRNOIQKQVQVOLIKKVGSKWEKDNLN-----EDEN 111

QY 109 SSNDASKVVEPKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKSG----- 154

QY 112 KKNGLISEQQV-----KEQOLRTITEQVKYQNLVFNMDYQOLDLNESGHRRHRETDY 164

Db 155 -INKLSQTSHLVLPSHAADGQTQVASFAFTPDKKTAIAEYTSRLGNGKPSRLIDQK 213

Db 165 DTEKWFELSH-----DQKMYVSIYA---NOKTSYCWMLDYFNKNYDHLNVSIN 211

QY 214 EIIDEGEIF---NAYQLTQLTIPNGYKSIGQDAFVDNK-----NIAEV- 253

Db 212 RLETEAEFYAFDDFSQTIKLT-NNSYQTVNTDVNFDDNNCTLLAIRFLLSLERFNILNIR 270

QY 254 -----NLPESELETISDYAFAHMSLKQVKLPPDNLKVIGELAFFFDNQIGGK--- 297

Db 271 SSYTRNQYNFEKIGELLETIFAWVFSHRHLOGIHLQVPCFAFOYLVNSSQISVKDSQLQ 330

QY 298 LY-----LPRHLIKAERAFAKSNRIQTVEFLGSKLKVGHEAS 334

Db 331 VYSFSTDILKLVDTNKVQDYFKFLOEFP-----LTHVSQLPVSATNAVNLNLKKVKA- 388

QY 335 FQDNNLIRNVMLPDGLEKIESEAFTGNPGDEHYNNQVLRTRGQNPOLATENTYVNPDK 394

Db 389 -----NLNLVSI-----TOFNFDFFYFVNLOHLKUEFGLEPNLTKQKL---E 428

QY 395 SIWRATPDMYTKWLEEDF-TY---QKNISVTGFSNKG--LQKVRNKNLEIPKQHNGITIT 449

Db 429 NLLLSIKQSKNLKFLRLNFYTYVAQETSRKQILKQATTIKNLKNNKNOEETPTKDETAPS 488

QY 450 EIGDNAFRNVDQFOSKTLRKYDLEEIKLPST---IRKIGAFAFQSNNLKSFEAS--EDLEE 504

Db 489 B-STSGMKFFDHLSELTELEDF-SVNLOQATEIYDSIHLKLIRSTNLKFKLSYKEMEK 546

QY 505 IKEGAF--MNRIGTL-DLKDKLLIKIGDAAFHINHIYAIVLPESV-----QEI--G 550

Db 547 SKMDTFLDLKNIYETLNNLKRCSVNISNP--HGNITISXELTNKDSTFYKFKLTNLQELQHA 604

QY 551 RSAFRONGALHLMFIGNKVKTIGEMAFLSNKLESV-----NLSEQKQLKTIEVQAFS 602

Db 603 DNALESEVVLPPNQLOTIREEAFKRNHL---KEVKGSSTLSQITFFNA-FDQNDGDKRFGKKV 658

QY 651 -NTIASLLYPNNTQ---KNPENPKPNLFFKQFQEQLKNLENVSINCTLDQ----- 695

Db 659 VVRTHNNSHMLADGERFIIDPPDKLSSSTMVDEKVLKIEGLDYSTLROTTQ----- 709

QY 696 -----HILNSISEFFLEKNKKIKAFILKRYYLQY--LDYTKLFKTLQQLPRLNQVY 745

Db 710 --TQFREMTTAGKALLSKSNLROGEGKQ-----FLOEAQFFIGRVDLDAIA---KA 756

QY 746 INQQLEELTVSE---VHKQWENHKQKAFYEPCLCEFKESSQTLQLIDFDQNTVSSDSI 801

Db 757 EKALVTKKATKNGHLLERSINKAVLAYNNSAIKKANVKRLEKELDLTDLVEKGKPLAQ 816

QY 802 KRILESISESKVHVLRLNPSQ---SSSLIKSEN---EEIQELLKACDEKGVLVKA 851

Db 817 TMVQGVVLLKTPPLPEYYIGLNVFD 843

QY 852 -----YKFPFLCLP----TGTYYD 866

RESULT 11

IF2_WOLSU STANDARD; PRT; 939 AA.

AC Q7M7X5; 15-MAR-2004 (Rel. 43, Created)
AC 15-MAR-2004 (Rel. 43, Last sequence update)
AC 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR WS2016.
OS Wolinella succinogenes.

NCBI_TaxID=844; [1]

SEQUENCE FROM N.A.

RETRAIN=DSMZ 1740;

RX MEDLINE=22882897; PubMed=14500908;

RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klimmek O., Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B., Meyer F., Lederer H., Schuster S.C.; "Complete genome sequence and analysis of Wolinella succinogenes." Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695 (2003).

RL CC - FUNCTION: One of the essential components for the initiation of protein synthesis. Protects formylmethionyl-tRNA from spontaneous hydrolysis and promotes its binding to the 30S ribosomal subunits. Also involved in the hydrolysis of GTP during the formation of the 70S ribosomal complex (By similarity).

CC -1- SUBCELLULAR LOCATION: Cyttoplasmic.

CC -1- SIMILARITY: Belongs to the IF-2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -

DR HAMAP; MF_00100; -; 1.

DR PROSITE; PS01176; IF2; 1.

KW Initiation factor; Protein biosynthesis; GTP-binding; Complete proteome.

FT DOMAIN 441 589 G-DOMAIN.

FT NP_BIND 447 454 GTP (BY SIMILARITY).

FT NP_BIND 493 497 GTP (BY SIMILARITY).

FT NP_BIND 547 550 GTP (BY SIMILARITY).

SQ SEQUENCE 939 AA; 102670 MW; B3D7BD9FB5CE3BC0 CRC64;

Query Match 3.7%; Score 200; DB 1; Length 939;

Best Local Similarity 19.7%; Pred. No. 0.034;

Matches 214; Conservative 148; Mismatches 365; Indels 360; Gaps 49;

QY 29 EREESVKQ---EQTQSASEDDWFE-EDNERKTNVSKENSTVDETVSDLFSDGNSSNNSSKT 85

Db 92 EVKESVKEAPESLPESPKEEAFAEALPKESVVKTPKLEQEPKKEELVSEPSLESASET 151

QY 86 ESVVSDPKQVKPKAKEPVTOEASNSSNDASKVVEPKQDTASKKETLETSTWEAKDFVTRGD 145

Db 152 --LSDSNPLPQEKE-TKETIVATLATQTDAEIQESEEKKETLAQATVQR----- 200

QY 146 TLVGFSSKGINKLSSOTSHLVLPSHAADGQTQVASFAFTPDKKTAIAEYTSRLGNGKP 205

Db 201 --VGLRI--VKKRSEE----PAPKADRPSLEE---ARTPSRTAGLKTQLSLGES-- 244

QY 206 SRLDIDQKEIIDEGEEIFNAYQLTK---LTIPNGYKSIGQDAFVDNKNIAEVNL-PES 258

Db 245 -----DESEAALARKKKEKKPLPAPT-----KNEQKIDLLGDR 281

QY 259 IETISDYAFAHMSLKQVKLPPDNLKVIGELAFFFDNQIGGKYLPRHLIKAERAFAKSNRIQ 318

Db 282 LETVSSF-----DDEQE---EIVLFDLTI-----RDDINKEDEVAK-KVD 317

QY 319 TVEFLGSKLKVGIGEASFQDNRLRNVMMLPDGLEKIESEAFTGNPGDEHYNNQVLRTRGQ 378

Db 318 T-----DRIVQRKTPFLDQGIRR-----KRRR 342

QY 379 NPHQLATENTYVNPDKSLWRAFPDMYTKWLEEDFTYQKNISVTGFSNKGLOKVRNNKLE 438

Db 343 RPQTVA-----DKESISCTIIP---EIRAYEFAEKTGKSIGEVIVKLFNLGLM 389

QY 439 IPKQHNGITITBEGDNAFRNVDQFOSKTLRKYDLEEIKLPSTIRKIG-AFAFQSNNLKSFE 497

Db	390	ITKN-----DFLDRDSTIELAEEFELLDVVVIKNTSEALEYSEE---E	428	RA
QY	498	ASEDLEEKEGAFMNNRIGTID---LKDKL---IKIGDAAFHINHIYAIVLPESVQR	548	RT
Db	429	EDEDEEGLEERPPVVITMGHVDHGKTSLLDKIRNTKVAAGEAGGITQHIGAYTVEKGK	488	RT
QY	549	IGRSAFRONGALHLMFIGNKVPIGEMAFLSNKLESVNLSSEQKQKTRIEVOAFSDNALS	608	RT
Db	489	I-----SFIDTPGHEAFTEMARGAEVTD---IVIIVIAADDGVKQ	526	RL
QY	609	VVLPPNLQQTREFAFKRNHLKEYKGSSTLSQTENAFFQNDGDKRFGKKVVVRTHNNHM	668	RN
Db	527	-----QTI---EA---LNHAKAANVPIITALNKVDKPD-----	554	[3]
QY	669	LADGERFIDPDKLSTMVLDKVLKLIEGDYSTLROTTQFREMTTAGKALLSKSNL	728	RP
Db	555	-----NPDKVKAEEAD-----LGYSPLEWGEYEFWHIS-----	583	SEQUENCE FROM N.A.
QY	729	ROGEKQKELQEAQFFLGRVLDLKAIKAKEAKLVTKKATKNGHLERSINKAVLAINNSAI	788	RX
Db	584	-----AKTGEGL-----DHLLE-----TILVQSBLLIE	605	Medline=95100974; PubMed=7802676;
QY	789	KKANVKRLEKELDLTLTDLVEKGKPLAQATMVQGVVLLKTPILPPIYYIGLNVYFDKSGKU	848	RA
Db	606	IKANPERKAKAVVISSSIEKGKGPVATVIVOSGTLKVGDSIVADTAYGRVRALIDDCGRN	665	RT
QY	849	IYALDMSD---TIGEGLQDAYGNPILNVDNEGYHTLAVATLADYEGLYI--KDLNS-	902	"Molecular cloning and sequence analysis of a human 372-kDa protein
Db	666	IQSIGPSEVAVVTGLSETPMAGAVLVSVENDS-----TAREYAEKRALYLRRQELSRST	719	localized in the Golgi complex.";
QY	903	--SLDKIKA-----IRQIP-LAKYHRLGIFTQAIRNA-----AAE	933	RT
Db	720	KVSFDELSAMVAEGQQLKSLPVVIKADTQGSLEKLRNEEVKINTIHAGVGGTE	779	!- SUBUNIT: Homodimer; disulfide-linked.
QY	934	ADRLILPKTPKGYL---NEVPNVRKKQMEKMLKPVDKYKPIFNKALPNEKVGDRAAKH	989	CC
Db	780	SDVVLAGASDNSVILGENVRPTGSVKRAKEL-GVEVKTYSIYAL---LDDVRAVLG	834	CC
QY	990	NINAETNNSAVTPIRSEQQLHKSQSDVNLIPOTSSKNNPYIYLGYVSLCLLFLVTAG-	1047	CC
Db	835	-----MMSPVLEEE-----NTGQAEVRETFIAYKVGTIAGC--LVTGSI	872	CC
QY	1048	KKGKRR 1054	CC	!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
Db	873	ORGIVKVR 879	CC	autoimmune disease Sjogren's syndrome.
RESULT	12	-----SIMILARITY: Belongs to the golgin family.	CC	
GOBL_HUMAN			CC	
ID	GOBL_HUMAN	STANDARD;	DR	
AC	Q14789;	014398; (Rel. 41, Created)	GO; GO:000139; C:Golgi membrane; TAS.	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	DR; GO:0016021; C:integral membrane; TAS.	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	DR; GO:0007030; P:Golgi organization and biogenesis; TAS.	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	DR; GO:0007030; P:Golgi organization and biogenesis; TAS.	
DE	(Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macro golgin) (Golgi complex-associated protein, 372-kDa) (GCP372).			
GN	GOLGB1.			
OS	Homo sapiens (Human).			
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	Medline=94187728; PubMed=7511208;			
RA	Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,			
RT	"Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (giantin)." Mol. Cell. Biol. 14:2564-2576 (1994).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	Medline=94257116; PubMed=8198703;			
RA	Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,			
QY	9	LAIALTTVSVVTYSQEVYG-LEREESVKQEOEQ-----SASEDDWFEEEDNERKTNVSKEN	62	Renz M.; "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as target of antibodies in patients with rheumatic diseases and HIV infections." J. Autoimmun. 7:67-91 (1994).
Db	326	LEVAERKLSFHNLQEEEMHILLEQFQAGQAOALESRYSALEQKHKAEEMEKTSHILSQ	385	
QY	63	STVDETIVS---DLFDGNNSSSSKTESVVSDPKVAKPEVTOEASNSSSNDASKVEVPK	120	
Db	386	KTGQELQASCDAKDQNSKLLQDNEQAVQSAQTIQLEQDQLQOKSKELSQFLNRLPLQQ	445	
QY	121	QDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSTSQSHMLPSHAADGTLQVA	180	
Db	446	HETA-----SOTS---FPDVNEGTO-----	463	
QY	181	SFAFTPDKKTAIAEYTSRIGENGKPSRL--DIDQKEITIDEGEIFNAYQVLTKLITPN	234	

Db 464 --AVTEENIASIQKRVVEL-ENEKGALLSIELEELKAENEKLSS-QITLLEAQNRTGE 519
 QY 235 GYKSIGQDAFDVNKNIAEVNLIPSELETSIDYAYA-HMSLKQVKLIPDNLKVIGELAFFDN 292
 Db 520 ADREVEISIIVDTIANKRSSAEESGQDVLENTFSQKHKL-SVLLLEMKEAQEEIAFLKL 578
 QY 293 QIGGKLYLPRHILKLAERA---FKSNRIOIVTEFLG---SKLKVIGEASFQDNLLRNVM 345
 Db 579 QLQG-----KRAEADHEVLDQKEMQMEGEGTAPIKMKVFLIEDTQDFPL---M 625
 QY 346 PD---GLEKTESEAFTGNPGDDEHNNQVVLRTTGTQ---NPHOLATENTYVNPDKSLR 398
 Db 626 PNEESSLPAVEKE---QASTER-----QSRITSEETISLNDAGVELKSTKQGDGKSL-S 673
 QY 399 ATPDMDYTKWLEEDFTYQKNSTVGFSNKGLOKVRRNKLEIPKQHNGITITEIGDNAFRN 458
 Db 674 AVPDIGQCH-QDELERLKSQII-----ELEL-NFHKAQEIYE-----KN 710
 QY 459 VDFQSKTLRKYD-LEEIKLPSTIRKIGAFAFQSNNLXSF-SED---LEEIKEGAFM 511
 Db 711 IDEKAKEISINLNQLIEEFKKNA-----DNNSAFTALSEERDOLLSQVKELSMV 759
 QY 512 NN---RIGTLDL---KDKLIKIGDAAFHINHIYATIVLPESEVOEIGSAFRONGALHL 562
 Db 760 TELRAQVKOLEMLAEEAERQRRLDYESQTAHDN---LLTEQI-----HS 800
 QY 563 MFIGNKVKTIGEMAFLSNKLESVNLSEQOKOLKTI-EVQAFSDNALSEWVLPPNLOTIREB 621
 Db 801 LSIEAKSKDV-KIEVLQNEELDDVOLQFSEOSTLIRSLQSOLONLKESEVL-----EG 850
 QY 622 AFKRNHLKRVKGSSTLSQITFNAFDQNDGDKRFRGKVVVRTHNNSHMLADGERFIIDPDK 681
 Db 851 AERVRHI-----SSKVEEL-----SQALSQKELEITKMDQ 880
 QY 682 LSSTMVDLKVLKIEGLDYSTIRQTQFREMTAGKALLSKSNLROGEKOKFLOEAQ 741
 Db 881 IL-----LEK-----KRDVETLQQTIEEKDQQV-----EISFSMTEKMVQLNEEK 921
 QY 742 FFLGRVDDIKATAKAEKALVTKATKNGHLLERSINKAVLAVNNSAIKA--NVKRLEK 798
 Db 922 FSLG-VEI-KTKEQNLNSRAEAKKEQVEEDNEVSSGLKQNYDEMSPAGOISKEELQH 979
 QY 799 ELDLILTDLVEKGKGPLAQATMVOGVYLLKTPPLPEYYIGLNVYFDKGKLIYALDMSDT- 857
 Db 980 EFDLILKKEENBORKRKLQQAALINRKLQRVSRLBEEANLK---DESKK---EIPLETH 1033
 QY 858 IGEGOKDAYGNPILNVDENEGCYHTLAVATLADYEGLYVIKDIILNSLDKIKAIROIPLAK 917
 Db 1034 RGEVEE-----DKENKEYSEKCVTCKQRIEIVLKQOTISEKEVELQHIRKDLEEK 1083
 QY 918 YHRLGIFQATRNAAAEADRLIPKTPKGYLNEPVNTYRKQOMEKNUKLPDVYKTPENKALPN 977
 Db 1084 L-----AAEEOFQOALVK---QMNQTLQDKTNQIDLLQAEISENQAIQKLITS 1128
 QY 978 --EKVDGDRRAAKGHNINAETNNSAVTP 1003
 Db 1129 NTDASDGDSVAL-----VKETTVVISP 1149

RESULT 13
 ID KTN1_HUMAN STANDARD; PRT; 1357 AA.
 AC Q86UP2; Q13999; Q14707; Q15387; Q86W57;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kinectin (Kinesin receptor) (CG-1 antigen).
 GN KTN1 OR CG1 OR KIAA0004.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TAXID=9606;
 RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymphoid;
 RX MEDLINE=95306853; PubMed=7787243;
 RA Fuettner A., Kruppa G., Kraemer B., Lemke H., Kroenke M.;
 RT "Molecular cloning and characterization of human kinectin.";
 RL Mol. Biol. Cell 6:161-170(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Peripheral blood lymphocytes;
 RX MEDLINE=94314220; PubMed=8039706;
 RA Print C.G., Leung E., Harrison J.E.B., Watson J.D., Kristiansen G.W.;
 RT "Cloning of a gene encoding a human leukocyte protein characterised by extensive heptad repeats.";
 RL Gene 144:221-228(1994).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Wang H.-C., Chen W.-F., Su Y.-R.;
 RT "Identification of a variant of *Homo sapiens* kinectin mRNA.";
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051387; PubMed=7584026;
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1.";
 RT DNA Res. 1:27-35(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX PubMed=12508121;
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
 RA Cattolico L., Levy M., Barbe V., De Bernardini V., Ureña-Vidal A.,
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cruaud C.,
 RA Brueis T., Jaillon O., Friedlander L., Samson G., Brottier P.,
 RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
 RA Aiach N., Boscub D., Dickhoff R., Dors M., Dubois I., Friedman C.,
 RA Gouyvenoux M., James R., Madan A., Mairey-Estrada B., Mangenot S.,
 RA Martins N., Menard M., Oztas S., Ratcliffe A., Snaffer T., Trask B.,
 RA Vacherie B., Bellemerre C., Belser C., Besnard-Gonnat M.,
 RA Bartol-Mavel D., Boutard M., Briez-Silla S., Combette S.,
 RA Dufosse-Laurient V., Ferron C., Lechaplais C., Louesse C., Muselet D.,
 RA Magdelanet G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
 RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,
 RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,
 RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
 RA Verdier J., Verdier-Discalca C., Hillier L.W., Fulton L., McPherson J.,
 RA Matsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,
 RA Quétier F., Waterston R., Hood L., Weissenbach J.;
 RT "The DNA sequence and analysis of human chromosome 14.";
 RL Nature 421:601-607(2003).
 RN [6]
 RP SEQUENCE OF 1-870 FROM N.A. (ISOFORMS 1/2).
 RX TISSUE=Lung;
 MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schueler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimmwood J., Schmutz J., Myers R.M.,

RA Kendrick K.E.;

RA Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.

RL [3] SEQUENCE OF 1-8 FROM N.A.

RA Bai Y.; Symington L.S.;

RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

Db -!- FUNCTION: Required for protein transport from the ER to the Golgi

Db -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR

Db MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE

Db ER AND THE GOLGI COMPLEX.

Db -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed

Db of an heptapeptide repeat pattern characteristic of alpha-helical

Db coiled coils. May form filamentous structures in the cell.

CC -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC DR EMBL; X54378; CAA38253.1; -

CC DR EMBL; L03188; AAB0143.1; -

CC DR EMBL; U53658; AAB66659.1; -

CC DR GermOnline; 140300; -

CC DR SGD; S0002216; USO1.

CC DR InterPro; IPR008938; ARM.

CC DR InterPro; IPR002017; Spectrin.

CC DR InterPro; IPR006955; USO1_p115_C.

CC DR InterPro; IPR006953; USO1_p115_head.

CC DR Pfam; PF04871; Us01_p115_C; 1.

CC DR Pfam; PF04869; Us01_p115_head; 1.

CC DR Transport; Protein_transport; Golgi_stack; Cytoskeleton; Coiled_coil.

CC FT DOMAIN 1 724 GLOBULAR_HEAD.

CC FT DOMAIN 725 1790 COILED_COIL<POTENTIAL>.

CC FT DOMAIN 465 487 CHARGED(HYPER-HYDROPHILIC).

CC FT DOMAIN 991 1790 DISPENSABLE_FOR_THE_PROTEIN_FUNCTION.

CC FT DOMAIN 1172 1786 ASP/GLU-RICH(ACIDIC).

CC FT CONFLICT 847 847 G->E (IN REF. 2).

CC FT CONFLICT 924 924 E->K (IN REF. 2).

CC FT CONFLICT 1253 1253 V->I (IN REF. 2).

CC FT CONFLICT 1319 1319 I->V (IN REF. 2).

CC FT CONFLICT 1461 1461 N->S (IN REF. 2).

CC FT CONFLICT 1581 1581 G->S (IN REF. 2).

CC FT CONFLICT 1600 1600 I->V (IN REF. 2).

CC FT CONFLICT 1661 1661 R->S (IN REF. 2).

CC FT CONFLICT 1772 1772 D->DEDDE (IN REF. 2).

CC SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

QY Query Match

QY Best Local Similarity 3.7%; Score 198.5; DB 1; Length 1790;

QY Matches 232; Conservative 209; Mismatches 432; Indels 329; Gaps 57;

QY

QY 16 VSVVTVSQEYVYGLERE-ESVKQEQTQASQEDDWFEED-NERKTKVSKENSTVDETVDLFE 73

QY

QY 724 ISKISF-EEVEKLRQCTKLGKETITSQETESTHENLTKLIALTNEHKELDEKYQIL- 781

QY

QY 74 SDGNSNNSSKTESVV--SDPKQVKAKPEVTQ----EASNSSNDASKVE---VPKQD 122

QY

QY 782 --NSSHSSLIKENFILETEIJKNVRDSDDEMTOIRDVLETKDKENQNTALLEYKSTIKQE 838

Db

Db 123 TASKK----ETLETSTWEAKDFVTR-GDTLVGFS-----KSGINKL 158

Db

Db 839 DSIKTLKGLETILSQKKKAEDGINKMGKDLFAULSREMQAVEENCKNQKEKDKSNVNHQ 898

QY "A cytoskeleton-related gene, us01, is required for intracellular

QY protein transport in *Saccharomyces cerevisiae*.";

QY J. Cell Biol. 113:245-260 (1991).

QY [2]

RP SEQUENCE OF 782-1790 FROM N.A.

RA Hostetter M.K., Herman D.J., Bendl C.M., McClellan M., Tao N.,

Db	936	KEHISKELVEYKSRFQSHDNLVAKLTEKLKSLANNYKDMQAENESLIKAVEESKNESIQ	995
Qy	254	--NLPESELETIS-----DXAFAMSLKQVKLPPDNLKVIIGELAFFDNQIGGK-----	297
Db	996	LSNLQNKIDSMSOEKENFOIERGSIKNEIQLK-----KTISDLEOTKEEISKSDDSKD	1050
Qy	298	-----LYLPRHLIKAERAFAKSNRIOVFELGSKLKVIGE-ASFQDNLRNMLPDGLEK	351
Db	1051	EYESQISLILKEKLETATTANDENTVNKISELTKTREELEAELAAYK--NLKN-ELETKLET	1107
Qy	352	TESEAFTGNPGDEHYNNOVVLTRTGQNPHQLATENTYVNPDKSLWRATPDMYTKWLEE	411
Db	1108	SEKALKEVKENEELKEEKI-----QLEKATETKQQLNSRA--NLESLEKEHE	1155
Qy	412	DFTYQ-KNSVTGFSNKGLOVKVRNNKLEIPKQHNGITITEIGDNAFR----NVDQSKTL	466
Db	1156	DLAAQLKKYEEQIANK----ERQYNEEISQLNDEITSTQQENESIKKKNDELEGEVKAM	1210
Qy	467	RKYDLEEIKLPSTIRKIGAFAFOSNNLKSFEASED---LEEIKEGAFMNNRIGTLDLKDK	523
Db	1211	KSTSEEQSNLKK--EIDALNLQIKELKKNETNEASLLESIK-----SVESE	1256
Qy	524	LIKIGDAAFHINHIYIAIVLPESVQEIGRSAFRQNGALHMFIGNKV--TIGEMAFLSNK	581
Db	1257	TVKIKELQDECNFKEKEV-SELEDKLKASEDKNSKYLELOKESEKIEELDAKTTTELKIQ	1315
Qy	582	LRSV-NLSEQOKQLKTIEVQAFS-----DNALSEVVLPPNLQTIREEAFKRNHLIKEVKG	633
Db	1316	LEKITNLSKAKEKESESELRSRIKKTTSSEERKNAEEQLEKLUKEIQKNOAFERFTIDPDKLSSTMV	687
Qy	634	SSTLSQITFNAFDQNDGDKRFGKKV-----VVRTHNNSHMLADGERFTIDPDKLSSTMV	687
Db	1376	SSTITQ-----EYSEKINTLEDELIRLQNEELKA-----KEIDNTRS	1413
Qy	688	DLEKVLKLIIEGLDYSTLROTTOTQFR-----EMTTAGKALLS--KSNLROGEKQ	734
Db	1414	ELEKV----SLSNDELLEEKONTIKSLQDEILSYKDKITRNDEKLSSIERDNKRDLES	1468
Qy	735	K-FLOEAQFFLGRVLDKAIKAKAEL--VTKKATKNGHLERS-----INKA	779
Db	1469	KEQLRAAQ-----ESKAKVEEGLKKLEEEESKEKELEKSKEMMKKLESTIESNET	1519
Qy	780	VLAYNNSAIKKA-----NVKRLEKE-LDLITDVLVEGKGPLAQATMVQGVYLL	825
Db	1520	ELKSSMETIRKSDEKLEQSKSAEEDIKLNQHEKSDLISRINESEKDIE-----L	1570
Qy	826	KTPLPLPEYYIGIINVFDKGKLIYALDMSDTIGEGOKDAYGNPILINVDEDNEGHTLAV	885
Db	1571	KSKLRIEA-----KSGSEI-----ETVKQELNNQAQKIRINAEE-----NTVLK	1609
Qy	886	ATLADYE-----GLYIKDILNSLDKIKAIROQIPLAKYHRLGIFOAIRNAAE	933
Db	1610	SKLEDIERELKDKQAEIKSNQEEKLLTSRLKELE-----QE	1646
Qy	934	ADRLLPKTPKGYLINEVPNVRKQMEKNLKPVDPYKTPIFNKPNEKVDGDRAGHNNINA	993
Db	1647	LDSTQQKAQKSEEERRAEVRKFQVEKS--QLEKAMLETKY-NDLVNKEQAWK-----	1697
Qy	994	ETNNNSVAVTPIRSEQOLHKSOSDV-NLPQTSKNNFIYELIGYVSLCLLFLVTAGKKR	1052
Db	1698	RDEDVTKTTDSQRQEIEKLAKELDNLKAENSKLKEANEDRSEIDDIMLLVTDLDEKNAK	1757
Qy	1053	AR 1054	
Db	1758	YR 1759	

Search completed: April 27, 2004, 11:15:52
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:

April 27, 2004, 11:11:11 ; Search time 27 Seconds

(without alignments)
3758.595 Million cell updates/sec

Title: US-10-091-007a-24

Perfect score: 5352

Sequence: 1 MTKKHLKTLALLTTVSVWT.....VSLCLLFLVTAGKKKARK 1055

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_78:
1: pir1:
2: pir2:
3: pir3:
4: pir4:
5: pir5:
6: pir6:
7: pir7:
8: pir8:
9: pir9:
10: pir10:
11: pir11:
12: pir12:
13: pir13:
14: pir14:
15: pir15:
16: pir16:
17: pir17:
18: pir18:
19: pir19:
20: pir20:
21: pir21:
22: pir22:
23: pir23:
24: pir24:
25: pir25:
26: pir26:
27: pir27:
28: pir28:
29: pir29:
30: pir30:
31: pir31:
32: pir32:
33: pir33:
34: pir34:
35: pir35:
36: pir36:
37: pir37:
38: pir38:
39: pir39:
40: pir40:
41: pir41:
42: pir42:
43: pir43:
44: pir44:
45: pir45:
46: pir46:
47: pir47:
48: pir48:
49: pir49:
50: pir50:
51: pir51:
52: pir52:
53: pir53:
54: pir54:
55: pir55:
56: pir56:
57: pir57:
58: pir58:
59: pir59:
60: pir60:
61: pir61:
62: pir62:
63: pir63:
64: pir64:
65: pir65:
66: pir66:
67: pir67:
68: pir68:
69: pir69:
70: pir70:
71: pir71:
72: pir72:
73: pir73:
74: pir74:
75: pir75:
76: pir76:
77: pir77:
78: pir78:
79: pir79:
80: pir80:
81: pir81:
82: pir82:
83: pir83:
84: pir84:
85: pir85:
86: pir86:
87: pir87:
88: pir88:
89: pir89:
90: pir90:
91: pir91:
92: pir92:
93: pir93:
94: pir94:
95: pir95:
96: pir96:
97: pir97:
98: pir98:
99: pir99:
100: pir100:
101: pir101:
102: pir102:
103: pir103:
104: pir104:
105: pir105:
106: pir106:
107: pir107:
108: pir108:
109: pir109:
110: pir110:
111: pir111:
112: pir112:
113: pir113:
114: pir114:
115: pir115:
116: pir116:
117: pir117:
118: pir118:
119: pir119:
120: pir120:
121: pir121:
122: pir122:
123: pir123:
124: pir124:
125: pir125:
126: pir126:
127: pir127:
128: pir128:
129: pir129:
130: pir130:
131: pir131:
132: pir132:
133: pir133:
134: pir134:
135: pir135:
136: pir136:
137: pir137:
138: pir138:
139: pir139:
140: pir140:
141: pir141:
142: pir142:
143: pir143:
144: pir144:
145: pir145:
146: pir146:
147: pir147:
148: pir148:
149: pir149:
150: pir150:
151: pir151:
152: pir152:
153: pir153:
154: pir154:
155: pir155:
156: pir156:
157: pir157:
158: pir158:
159: pir159:
160: pir160:
161: pir161:
162: pir162:
163: pir163:
164: pir164:
165: pir165:
166: pir166:
167: pir167:
168: pir168:
169: pir169:
170: pir170:
171: pir171:
172: pir172:
173: pir173:
174: pir174:
175: pir175:
176: pir176:
177: pir177:
178: pir178:
179: pir179:
180: pir180:
181: pir181:
182: pir182:
183: pir183:
184: pir184:
185: pir185:
186: pir186:
187: pir187:
188: pir188:
189: pir189:
190: pir190:
191: pir191:
192: pir192:
193: pir193:
194: pir194:
195: pir195:
196: pir196:
197: pir197:
198: pir198:
199: pir199:
200: pir200:
201: pir201:
202: pir202:
203: pir203:
204: pir204:
205: pir205:
206: pir206:
207: pir207:
208: pir208:
209: pir209:
210: pir210:
211: pir211:
212: pir212:
213: pir213:
214: pir214:
215: pir215:
216: pir216:
217: pir217:
218: pir218:
219: pir219:
220: pir220:
221: pir221:
222: pir222:
223: pir223:
224: pir224:
225: pir225:
226: pir226:
227: pir227:
228: pir228:
229: pir229:
230: pir230:
231: pir231:
232: pir232:
233: pir233:
234: pir234:
235: pir235:
236: pir236:
237: pir237:
238: pir238:
239: pir239:
240: pir240:
241: pir241:
242: pir242:
243: pir243:
244: pir244:
245: pir245:
246: pir246:
247: pir247:
248: pir248:
249: pir249:
250: pir250:
251: pir251:
252: pir252:
253: pir253:
254: pir254:
255: pir255:
256: pir256:
257: pir257:
258: pir258:
259: pir259:
260: pir260:
261: pir261:
262: pir262:
263: pir263:
264: pir264:
265: pir265:
266: pir266:
267: pir267:
268: pir268:
269: pir269:
270: pir270:
271: pir271:
272: pir272:
273: pir273:
274: pir274:
275: pir275:
276: pir276:
277: pir277:
278: pir278:
279: pir279:
280: pir280:
281: pir281:
282: pir282:
283: pir283:
284: pir284:
285: pir285:
286: pir286:
287: pir287:
288: pir288:
289: pir289:
290: pir290:
291: pir291:
292: pir292:
293: pir293:
294: pir294:
295: pir295:
296: pir296:
297: pir297:
298: pir298:
299: pir299:
300: pir300:
301: pir301:
302: pir302:
303: pir303:
304: pir304:
305: pir305:
306: pir306:
307: pir307:
308: pir308:
309: pir309:
310: pir310:
311: pir311:
312: pir312:
313: pir313:
314: pir314:
315: pir315:
316: pir316:
317: pir317:
318: pir318:
319: pir319:
320: pir320:
321: pir321:
322: pir322:
323: pir323:
324: pir324:
325: pir325:
326: pir326:
327: pir327:
328: pir328:
329: pir329:
330: pir330:
331: pir331:
332: pir332:
333: pir333:
334: pir334:
335: pir335:
336: pir336:
337: pir337:
338: pir338:
339: pir339:
340: pir340:
341: pir341:
342: pir342:
343: pir343:
344: pir344:
345: pir345:
346: pir346:
347: pir347:
348: pir348:
349: pir349:
350: pir350:
351: pir351:
352: pir352:
353: pir353:
354: pir354:
355: pir355:
356: pir356:
357: pir357:
358: pir358:
359: pir359:
360: pir360:
361: pir361:
362: pir362:
363: pir363:
364: pir364:
365: pir365:
366: pir366:
367: pir367:
368: pir368:
369: pir369:
370: pir370:
371: pir371:
372: pir372:
373: pir373:
374: pir374:
375: pir375:
376: pir376:
377: pir377:
378: pir378:
379: pir379:
380: pir380:
381: pir381:
382: pir382:
383: pir383:
384: pir384:
385: pir385:
386: pir386:
387: pir387:
388: pir388:
389: pir389:
390: pir390:
391: pir391:
392: pir392:
393: pir393:
394: pir394:
395: pir395:
396: pir396:
397: pir397:
398: pir398:
399: pir399:
400: pir400:
401: pir401:
402: pir402:
403: pir403:
404: pir404:
405: pir405:
406: pir406:
407: pir407:
408: pir408:
409: pir409:
410: pir410:
411: pir411:
412: pir412:
413: pir413:
414: pir414:
415: pir415:
416: pir416:
417: pir417:
418: pir418:
419: pir419:
420: pir420:
421: pir421:
422: pir422:
423: pir423:
424: pir424:
425: pir425:
426: pir426:
427: pir427:
428: pir428:
429: pir429:
430: pir430:
431: pir431:
432: pir432:
433: pir433:
434: pir434:
435: pir435:
436: pir436:
437: pir437:
438: pir438:
439: pir439:
440: pir440:
441: pir441:
442: pir442:
443: pir443:
444: pir444:
445: pir445:
446: pir446:
447: pir447:
448: pir448:
449: pir449:
450: pir450:
451: pir451:
452: pir452:
453: pir453:
454: pir454:
455: pir455:
456: pir456:
457: pir457:
458: pir458:
459: pir459:
460: pir460:
461: pir461:
462: pir462:
463: pir463:
464: pir464:
465: pir465:
466: pir466:
467: pir467:
468: pir468:
469: pir469:
470: pir470:
471: pir471:
472: pir472:
473: pir473:
474: pir474:
475: pir475:
476: pir476:
477: pir477:
478: pir478:
479: pir479:
480: pir480:
481: pir481:
482: pir482:
483: pir483:
484: pir484:
485: pir485:
486: pir486:
487: pir487:
488: pir488:
489: pir489:
490: pir490:
491: pir491:
492: pir492:
493: pir493:
494: pir494:
495: pir495:
496: pir496:
497: pir497:
498: pir498:
499: pir499:
500: pir500:
501: pir501:
502: pir502:
503: pir503:
504: pir504:
505: pir505:
506: pir506:
507: pir507:
508: pir508:
509: pir509:
510: pir510:
511: pir511:
512: pir512:
513: pir513:
514: pir514:
515: pir515:
516: pir516:
517: pir517:
518: pir518:
519: pir519:
520: pir520:
521: pir521:
522: pir522:
523: pir523:
524: pir524:
525: pir525:
526: pir526:
527: pir527:
528: pir528:
529: pir529:
530: pir530:
531: pir531:
532: pir532:
533: pir533:
534: pir534:
535: pir535:
536: pir536:
537: pir537:
538: pir538:
539: pir539:
540: pir540:
541: pir541:
542: pir542:
543: pir543:
544: pir544:
545: pir545:
546: pir546:
547: pir547:
548: pir548:
549: pir549:
550: pir550:
551: pir551:
552: pir552:
553: pir553:
554: pir554:
555: pir555:
556: pir556:
557: pir557:
558: pir558:
559: pir559:
560: pir560:
561: pir561:
562: pir562:
563: pir563:
564: pir564:
565: pir565:
566: pir566:
567: pir567:
568: pir568:
569: pir569:
570: pir570:
571: pir571:
572: pir572:
573: pir573:
574: pir574:
575: pir575:
576: pir576:
577: pir577:
578: pir578:
579: pir579:
580: pir580:
581: pir581:
582: pir582:
583: pir583:
584: pir584:
585: pir585:
586: pir586:
587: pir587:
588: pir588:
589: pir589:
590: pir590:
591: pir591:
592: pir592:
593: pir593:
594: pir594:
595: pir595:
596: pir596:
597: pir597:
598: pir598:
599: pir599:
600: pir600:
601: pir601:
602: pir602:
603: pir603:
604: pir604:
605: pir605:
606: pir606:
607: pir607:
608: pir608:
609: pir609:
610: pir610:
611: pir611:
612: pir612:
613: pir613:
614: pir614:
615: pir615:
616: pir616:
617: pir617:
618: pir618:
619: pir619:
620: pir620:
621: pir621:
622: pir622:
623: pir623:
624: pir624:
625: pir625:
626: pir626:
627: pir627:
628: pir628:
629: pir629:
630: pir630:
631: pir631:
632: pir632:
633: pir633:
634: pir634:
635: pir635:
636: pir636:
637: pir637:
638: pir638:
639: pir639:
640: pir640:
641: pir641:
642: pir642:
643: pir643:
644: pir644:
6

A;Molecule type: DNA
 A;Residues: 1-1569, 'V', 1571-1821, 'K', 1823-1959, 'RHL', 1963-1970, 'RNP', 1973-2048, 'N', 2050-2
 A;Cross-references: EMBL:X61236; NID:94071; PID:CAA43554.1; PID:94072
 C;Genetics:
 A;Gene: SGD:NUM1
 A;Cross-references: SGD:S0002557; MIPS:YDR150w
 A;Map position: 4R

Query Match 4.2%; Score 224; DB 2; Length 2748;
 Best Local Similarity 20.0%; Pred. No. 0.035; Gaps 50;
 Matches 228; Conservative 191; Mismatches 464; Indels 258; Gaps 50;

QY 7 KTLALALATTVSVVTYSQEVYGLEREEVKQBOTQASASDDWFEEDNERKTNVSKENSTV 66
 Db 193 ESLKLKLKDLEILTLTKN-GMENDLS---SOKLHYDEKISELKERILDNNENDRLI 245

QY 67 ETVSDLFSGNSNNSSKTESVVDPKQVKAKPEVTO-----EASNNSNDASKVE 117
 Db 246 ISVSDLTSEINS-LQSNRTER-IKIQKOLDKASISLKRKVQKYYQKQHTSDTIVTS 303

QY 118 VPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSPAHDGTOLT 177
 Db 304 DPDSSEGTTSEEDIRDIVIRIDHMIETGPSWEDISEDLVKKYSEKNNMILLSNSDYSKNUQ 363

QY 178 QVASFAFTPDKKTAIAEYTSRIGENGKPSRDLIDQKETIDEGEIFFNAYQLTQTLTIPNGYK 237
 Db 364 KSESASKPKODELMTKEVAENLNMIALPNDNYSKKEFSLESHI-----KYLEASGYK 416

QY 238 SIGQDAFVDNKNTIAEVN-----LPESLETT----- 262
 Db 417 VLPLERPF-ENLNESLSNPSVSYNLKEKLQALKKIPIDQSTFNLLIKEPTIDFLPLTSKDC 475

QY 263 -----SDY-----AFAHMSLKQVK-----LPDNL-----KVIGELAFFD--N 292
 Db 476 LIIPTKDXNDLFESVKVNPNSIEQMKKCLEARNLQSNICKWLEERNNGCKWLSNDLYFSMVN 535

QY 293 QI-GGKLYLPRHL-----IKLAERAFKSNRIQTVFLLGSKLKVIGEASFQDNNL--RNV 343

Db 536 KIETPSKQYLSDKAKEYDQVLIDTKALEGLKKNPTIDFLREK-----ASASDYLILKED 589

QY 344 MLPDGLEKIESEAFTGNP--GDEHYNNQV-----VLRTRTGQNPQHOLATENTYVN 391

Db 590 YVSPSLEYLVEHAKATNHHLLSDSAYEDLVKCKENPDMEFLIKEKSAKLGHTVSNEAYSE 649

QY 392 PDKSLWRAATPDMDYTKWLEEDFTYQOKNSVTGFSNKGQVKVRNNKOLEIPKQHNGITITEI 451

Db 650 LEKKL--EOPSLERYL--VEHAKATNHHLLSDSAYEDLVKCKENPDMEFLKERS---AKL 701

QY 452 GDNARFNWDQFSKTLRKY-DLE-EIKLPSTIRKIGAFAFQSNNLKSFEASEDLEBIKEGA 509

Db 702 GHTVVSN-EAYSELQRKYSELEKRVQPSLVLVLEHAKATDHHLLSDSAYEDLVKCKENP 760

QY 510 FMNNRIGTLDLKDQLIKGDAAFHINHIYIAIVLPESVQEIGRSAFRONGALHLMFIGNKV 569

Db 761 DVE-----FLKEKSAKLG-----HTVVSSEEEYSELQR----- 787

QY 570 KTIGEMAFLSNKLESVNLS-EQKLTIEVQAFSDNALSEVVL---PPNLQTIREEAFKR 625

Db 788 ---KYELEKEVEQPSLAVLVEHAKATDHHLLSDSAYEEVLLVCKKENPDMEFLKEKSAKL 843

QY 626 NHLKEVKGESTLSQTTFNAFDONDGDKRGKVVVRTHNNSHMLADGERFLIDPDKLST 685

Db 844 GH-----TVVSNEAYSELEKLEQPSLAVLVEHAKATDHHLLSDSAYEDLVKCKENSD 896

QY 686 MVDL-EKVLK---TIEGLDYSTL-ROTTOTFREMTTAGKA---LLSKS-----NLR 729

Db 897 VEPFLKEKSAKLGLHTVVSNEAYSELEKLEQPSLAVLVEHAKATDHHLLSDSAYEDLVKCK 956

QY 730 QGEKOKFQLEAQOFFIGR-VDLDDKAIAKAKA-----LYTKKATKNGHLL----- 772

Db 957 ENPDMEFLIKEKSAKLGLHTVVSNEAYSELEKLEQPSLAVLVEHAKATNHHLLSDSAYEDL 1016

QY 773 -----ERSINKAVLAVNNSATKANKVRIEKELDLTDLVEGKGP--LAQAT 817

RESULT 4

B42771 reticulocyte-binding protein 2 - Plasmodium vivax (fragment)

C;Species: Plasmodium vivax

C;Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Sep-1997

C;Accession: B42771

R;Galinski, M.R.; Medina, C.C.; Ingravollo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992

A;Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.

A;Reference number: A42771; MUID:92315338; PMID:1617731

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1252 <GAL>

A;Cross-references: GB:MB88098; NID:9160627; PID:9160628

A;Experimental source: strain Belem, merozoites

C;Genetics:

A;Gene: RBP2

Query Match 4.2%; Score 223.5; DB 2; Length 1252;

Best Local Similarity 18.4%; Pred. No. 0.011; Gaps 52;

Matches 219; Conservative 190; Mismatches 414; Indels 369; Gaps 52;

QY 11 LALTTVSVVTYSQEVYGLEREEVKQEQTOS---ASEDWFEEDNERKTNVSKENSTV- 65
 Db 126 IATTKKGTSDTSQDINEL--ESIKEEVHKNLQIYQKQESNSMEE--MRKQILSMKDLIL 180

QY 66 --DETVSDLFSDG-----NSNNSSSKTBSVVS-----DPKQVPKAKPEVTOEASN 108

Db 181 NNSETIAKEISNNNTQNALGFRENRAKTKLNKTDLILQRAVAMIEEAKHKNNDIALEDAQ 240

QY 109 SSNDASKVKEVKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVPS 168

Db 241 IDETEVKIEQINREIMKKDEIKSYLSEIKEYDKCTTEISNSKRGKDKIEFLE----- 294

QY 169 HAADGTOLQVQASFTPDK-----KTAIAEYTSRIGENGKPSRLLID--OKEITIDEGET 221

Db 295 -----KFPNBEESNSNKVNINENENI-RNSEQYLKQDIEDAEKOASTKVEL 339

QY 222 FNAYQLTQTKLIPNGYKSTIGQDAFVDNKNTIAEVNLPESLETTSDYAAHMSLKVQVLPDNL 281

Db 340 FHKHE--TTISNIFKE-SEILGVETKSQKINKAEDIMKEIERHNSEIQTQVKGFFQENL 395

QY 282 KVIGELAFFDNOIGGKLYLPRHLKLAERAFKSNRIQTVFLLGSKLKVIGEASFQD--NN 339

Db 396 NKLNEPHNYDN-----AEDELNNDK-----STNAKVLIETNJESVHN 433

QY 340 LRNVM-LPDGLEKIESEAFTGNPGDEHYNNQVLRTRTGQNPQHOLATENTYVNPDKSLWR 398

Db 434 LSEITNIKQGGEKJYSA-----KDIMQKIK-----TSENT---AEKTEK 472

QY 399 ATPDM-DYTKWLEEDFT-----YQKNSVTGFSN-----KGLQVKVRNKMLEIPKQHNGI 446

Db	473 VKDDQSNTVNLNQITTERNLIVTEKNRINGIDSTITNIEGALKESKG-NYEI-----GF	526	Qy	3 KKHKLTLALLTIVSVVTYSQEVYGLEREESVKQEQTOSASEDDWFEEEDNERKTNVSKEN	62
Qy	447 --TITEIGDNAFRNVDFOQSKTLRKYDLEEKLPLSTIRKIGAFAFQSN--LKSFEASED	502	Db	883 KREAENALASAATAKTGELEAKIQTDLDEDKISELESKLSAELDK--QBLNLKTIENLEEDK	940
Db	527 LEKLBEBIGKNRKLKDVLTKSINS-----		Qy	63 STVDETVDLFLSDGNSN-----NSSKT-----	
Qy	503 E--EIKEGAFMNNRIGTDLKDKLIKIGDAFHINHYATIVLPESVOEIGRSAFRONGAL	560	Db	941 EELKETIDNLKGDLDSKLKGDEDLEVEITELNSQINTLNATVDKKDTTAEMQESIDEKE	1000
Db	574 NDYENKMGETYNEFEGSLN-----		Qy	93 KQVPKAKPEV--TOEASSNSNDASKVEVPKQDTASKETLETSTWAKDFVTRGDTLV-	148
Qy	561 HLMFIGNKVKTIGEMAFLSNKLESVNLSEQKOLKTIEVOAFSDNALSEVVLPPNLIQTRB	620	Db	1001 DETKLKGDLKLEEKDDLEQDRAVSATKDIACKLNKITECEDAKDETAKELE	1060
Db	602 -----SENTSDYNSAKTLRLEAQKEV-----		Qy	149 --GFSKSGINIKLSQTSHLVLPSHAADGTTQASFAFTPDKTAIE-----YTSR	198
Db	621 EAFKRNLKEVKGSSTLSQITFNAFDQNDGDKRGK--VVTRTHNNSHMLADGERFI	677	Db	1061 DEENKNKDLTNELOQTO--LKUGETEKSLSAAQVATKASDERDTISOLNEKLTTKN	1117
Qy	631 EANK--YLRDVKVKVESF-RFIEFNKESLDKINEMIKKEQLTNEGHGNVKQLIVENIKELV	687	Qy	199 LGE----NGKPSRLIDQKELIID----EGHIFNAY----OLTKLTIPNGY	236
Db	678 DPDKLSSSTMVDLEKVLKIEGLDYSTLROTTQFREMTAGK-----ALLS	724	Db	1118 LTKTKADLEKKISGLKQDYDLEDDKNCIEGDLRNAQORKIKELDDEITKGADVSQYILOQO	1177
Db	688 DENNLSDTILKQATGKNEEIQKITHSTLKNKAKTILGHVDTSAKYVGIKITPELALTELLG	747	Qy	237 -----KSIGQDAFDVNKNIAE----VNLPESI-ETISDYAFAHMSLKV	275
Qy	725 KSNLROGEKQKFLQBAQFFL-----GRVLDKAIKAKEAL-----VTKKAT	766	Db	1178 KEYESQIAKMQEEKEAIGNDVKNKEKTIKELEIQSLSQEKDETVEKEKKEI	1237
Db	748 DAKLKTAQELKFESKMNVLETEENMSKNTNELDVHKNTIDDAYKVALEILAHSDDEIDTKQK	807	Qy	767 KNHILLERSIN---KAVL--AYNS-----AIKKANVKRLEKELDLTDLVEGK	810
Qy	808 DSSKLIEMGNQIYLKVLINQYKNKLISSIKSKEEAVSVKIGNVSKKHSELSKIT	861	Db	1238 EKEMKALQEKEKENVESSKMNSTEKKKKLEDNLIKDTQKKLDDMT-ADNE-----	1284
Db	862 -----CSDKSYDNTIALEKTEL-QNDRNSFTQEK	890	Qy	276 -----KLPDNL----KVIGELAFFDNQIGGKLYLPRHLI	305
Qy	871 LNVDEDN-----EGYHTLAVATIADVEGLYK-----	897	Db	1285 KLIKAKA-KDLEAQOLNEWQDNHEHKAVADAEELINKKAQSDKELNSLXAELEALTAKSV	1343
Db	891 TNTNSDSKULEKIKTDFESLKNALKTLEGEVNALKASSDNHEHQVQSKSEPVPNALSETEKE	950	Qy	358 TGNPGDEHYNQVVLRTTGTQNPOLATENTYNPDKSLWRATPDMD-----	404
Qy	898 ---DILNSSLDK-IAIRQIPLAKYHRLGIFOAIRMA-----ABADRLLPKTPKGY	945	Db	1344 SKNKDSE--NEKAALSEHIDQANEKL-----NIQADLRKATADLQEANEKKAIVEAQ	1395
Db	951 ETDIDSNTALDELLKKGRTCBEVSRY-----KLIKDTVTKEISDDTELINTIEKVNKAY	1004	Qy	405 -----YTKWLE-----EDFTY-----QKNSVTGFSNKGL--QKVRRNK	436
Qy	946 LNEVPNRYKKQOMEKONLKPVDYKTPIFN-KALPN-EKVDGDRRAAKHN--INAETNNSAV	1001	Db	1396 DKLWADNKKWTKTLEEIKARDEENTYKVENXEKVLKRKEADEEANENLIDIEKDRMNKE	1455
Db	1005 L---AYIKKNEYDTVQDVLTLINEHFNTKQVSNSHEPTNFDSKSNSEELTKAVTDSKTI	1060	Qy	437 LEIPKQHNGTITEIGDNAF--RNVDFOQSKTLRKYDLEI-----KLPSTIR	481
Qy	1002 TPIRSEQQLHKSQSDVNLQPTSKN-----NFIYETIGYVSL	1038	Db	1456 KQVKKLEGELEKETKDKLNAAIAEKSIFTAKQSDADLEELINKTVEHDEVVARLNTQIT	1515
Db	1061 SKLKGVIIEVNENTEMNTIESSAKEIEALYNELKONKTSLSNEIYQTSNEVKL	1112	Qy	482 KI---GAFAFQSNLKSFEASEDLEEIKE---GAFMNNRIGTLDLKDQLKLIKIGDAF	532
RESULT 5			Db	1516 KLTDRNQSAEEELNRS-KADKKKISELLEQVNLESRPYGTGNADENEIKRIDA--	1572
T18296	myosin heavy chain - Entamoeba histolytica		Qy	533 HINHIIYATIVLPESVOEIGRSAFRONGALHLMFIGNKVK-TIGEMAFLSN-----	580
C;Species:	Entamoeba histolytica		Db	1573 QIADLINKALEMKGVQNNQLOATNKLAKNDLTSKIEITENEMKLENAKRLEQDKE	1632
C;Accession:	T18296 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000		Qy	581 -----KLESVNLSSEQKOLKTIEVOA-FSDNALSEWV-----LPPNLQQTIREA	623
R;Guillen, N.	submitted to the EMBL Data Library, February 1997		Db	1633 ADKAVSEQTIRKKGLEEVKKLTTIQLQALKFOINAPSSVQEEKQRLESDAELKEQLE	1692
A;Reference number:	Z18865		Qy	624 KRN-----HLKEVGSSTLSQITFNAFDQNDGDKRGKVVVRTHNNSH-----	668
A;Accession:	T18296		Db	1693 QERTTAANAEERKKQIAELDEVKFNLEDVTNQRE--KLVAKNSENDAEIDSIKEEKKA	1749
A;Status:	preliminary; translated from GB/EMBL/DDBJ		Qy	669 LADG-ERFIDPDKLSSSTMVDLEKVLKIEGLDYSTLROTTQFREMTAGKALL--SK	725
A;Molecule type:	DNA		Db	1750 LEDELEKITTDDNNKLSSEEDSLDRKYNAL--LDSKSDVSMKEKQFDELKVTKOALETEK	1807
A;Residues:	1-2139 <GUT>		A;Cross-references:	EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1	
A;Genetics:			A;Gene:	mhca	
C;Superfamily:	myosin heavy chain; myosin motor domain homology <MMO>		F;91-780/Domain:	myosin motor domain homology	
Query Match	4.0%; Score 216; DB 2; Length 2139;		Best Local Similarity	20.2%; Pred. No. 0.056; Mismatches 494; Indels 342; Gaps 58;	
Matches	260; Conservative 188; Mismatches 1864; NELD-----DVKDQLKAQDDIADKEDELATLDQKXKTLVQKSVFDSRIQEMQEQLD		1916		

Query Match 4.0%; Score 212.5; DB 2; Length 621;
Best Local Similarity 21.9%; Pred. No. 0.013; Mismatches 107; Conservative 73; Indels 159; Gaps 20;

RESULT 6

QY 838 INVYFDKSGKLIYALDMDSDTIGEGQKDAYGNPILNV---DEDNEYHTLA---VATLADY 891
Db 1917 L---EKAGR-----AKAQKOKOAYEKKQLOELOENDNDFEETYKETADKINTLSAQ 1963
QY 892 EGLYIKDILNSLDKIKAIROIPLAKYHRLGI-FOATRNAAE---ADRLLPKTPKGYL 947
Db 1964 -----KDDLQKELEKERGLRQDSEKEVORLVRVKCQELETKVAEVGGANVIAKVKAYEA 2018
QY 948 EVPNY-----RKKOMEKNIKPVDYKTPIFNKLALPNEKVGDGRRAAKGHNIAETNNS 998
Db 2019 EIEELTTEAEDALKAKMKAEKKAKTSOKKDELQKTI-----ADYETKEASFNTIEIG-- 2070
QY 999 VAVTPIRSHQQLHKSQSDVNLPQT 1022
Db 2071 -----KTQAEKKYQQVRDDET 2088

Query Match 4.0%; Score 215.5; DB 2; Length 1727;
Best Local Similarity 19.3%; Pred. No. 0.043; Mismatches 422; Indels 343; Gaps 57;
Matches 231; Conservative 198; Mismatches 422; Indels 343; Gaps 57;

QY 12 ALTTVSVVTVSQEVYGLERBESVKQEQOTQASED----DWFERDNERKTNVSKENSTV 66
Db 116 ALTT-----ENQSLRANSLEQOSKIAQDQTEALQENENSHLGEQVQSAH 167
QY 67 ETVSD-----LFDGNSNNSSKTESVSDPKQVKAKPVTQEAQNSNDASKVEV 118
Db 168 QALSDIEERKKQHMF---SSSRVKEELIVOE-----KSALMVSBLASLQSDHSKV-- 215
QY 119 PKQDTASKETLETSTWEAKDFVTRGDTLVGFSKSG---INKLSQTSHLVLPSHAADG-- 173
Db 216 -----CEKLEVSRSQVQDLEKK---LAGLQAQNTTELNEKIQLEQKRSNQYSSDGNIS 264
QY 174 -----TOLTOVASFAFTPDKTKTAIAEYTSRLGENGKPSRLDID---QKEIIIDEGEIF 222
Db 265 KILETDPTSIKELEEEVETOKRITALWE-----SKSSELQSEVAALQEKLTQSQSLY 316
QY 223 NAYQLTK-----LTPNGYKSIGQ-DAFVDNKNTIAEVN-----LPE 257
Db 317 N-NVTEELNNKQQLIISENSIRELQHKYDSVSELQVKENKNTVSAGVGLFSPIAQ 374
QY 258 SLETIS---DYFAFAHMSLKVOKVLPDNL-----KVIGELAFFFDDNQIGGKLYLPRHLIKL 307
Db 375 KLSAVQNPESFSFTKVISDNMKLQQVKVSSLKLQDLRDLTINKFSSFCEQV--KQRIPI--VVKQ 430
QY 308 AERAFKSNRIQLOTEFVGSKLKVIGEASFQDNNLRNVMPLPDGLEKIESEAFTGNPGDEHYN 367
Db 431 QRSERIVRNNTY-MNFLESLET-----SNNLTKVQAE LISTKMRQEA-----YL 475
QY 368 NQVVLRLTRTGONPHQ---LATENTYVNPDKSLWLRATPDMYTKMLEEDFTYQKNSVTG-- 422
Db 476 QLTASRTOCSDLRSREVICLMAELDHINETKS--RNVPATVQVA-LDE--YAQNPTASE 529
QY 423 -FSNKGLQ-----KVRRNKNLEIPQHNGITITEIGDNAFRNVDQSKTLRYDLEEIKL 476

Query Match 4.0%; Score 215.5; DB 2; Length 1727;
Best Local Similarity 19.3%; Pred. No. 0.043; Mismatches 422; Indels 343; Gaps 57;

RESULT 7

QY 95250 choline binding protein *pepa* [imported] - *Streptococcus pneumoniae* (strain TIGR4)
C;Species: *Streptococcus pneumoniae*
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: A95250
R;Tettelin, H.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heisler, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, T.; Hickey, E.K.; Holt, I.B.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, T.; Hickey, E.K.; Holt, I.B.
A;Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: A95250
A;Status: preliminary
A;Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
A;Molecule type: DNA
A;Residues: 1-621 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK76194.1; PID:g14973649; GSPDB:GN00164; TIGR:S
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: Sp2136

Query Match 4.0%; Score 212.5; DB 2; Length 621;
Best Local Similarity 21.9%; Pred. No. 0.013; Mismatches 107; Conservative 73; Indels 149; Gaps 159; Gaps 20;

Db 530 TLYNICKELANFSSSIKEAVSKTELREREKALEC-----DVEIOKOTVOXOISNAVKE 580
QY 477 PSTIRKIGAFAFQSNNLKSFEASEDLERIKEGAFMNRR-----IG 516
Db 581 NSNT-----LSEQITNLESELNNSKTKNEELNERNILKEMLATSRSSTLHNSSAG 632
QY 517 TLDLKOKLIKIGDAAFHINHIAIYIVLPE--SVOEIGRSAFRONGALHLMFIGNKVKTIGE 574
Db 633 NID--DKMKSIDESTRELEKNYEVRNEMTAQE--SLSKRNQDL-----LSE 676
QY 575 MARLSNKLESVNLSSEQQLKT-----IEVOQAFSDNALESEVVLPPNPLQTI--REEAFKR 625
Db 677 MEAIRKELE--NSKVQQQLSTDRLTMANNDVEAFKKEAKELRSINQNLQDIISRQDQRAS 734
QY 626 NHILKEVKGSSTLSQITENAFDONDGDKRGKVVVRTHNNSHMLADGERFTIDPDKLST 685
Db 735 KFAEELLHVNLSLAERLKGEELNASKGEKDLRKRT-----QERLISENDKL-- 778
QY 686 MVDLDEKVVKITIEGDYSTLRTQTOTQFREMTTAGKALLSKSNLROGEKOKFLOEAQFFIG 745
Db 779 LAERERLMSLIVSDL-----QTFLMQ-QQLSDAAR-----KVFESEKE--- 815
QY 746 RVDLDKALAKAEKALVTKKATKNGHILERSINKAVLAYNN-----SAIKKANVRL 796
Db 816 --SLSLSLQKLIKES--NEKMSNDLHSLOKSLEKSGIEYSSRIKTLMBKQSLSEDNRKJL 871
QY 797 E-----KELDLITDVEKGKGLPQAQTMVQ-----GVYLLKTPL-PLPEYYIGL 838
Db 872 DNQQMMIEKIKLQBLINGVIELEKQRFSTLEAKFTQOKNTSYSEREALLESSLSLQSKHTSL 931
QY 839 NWYFDKSGKLIYALDMDSDTIGEGQ---KDAYGNPILNVDENDEGYHTLAVATLADYEGL 894
Db 932 ESDYDYNYSIRNTEQLOQAAASKLAEEMVERVKTTEYDEYRQIOTSESLEKHN-LKITSLEQRIVI 990
QY 895 YKDILNSLDKIKAIROIPLAKYHRLGIFQAIRNAAEAD--RLLPKTPKGYLNEVFN 951
Db 991 LQDEIASSL-----RCENITKDSETRVALLIEENKHNLSS 1029
QY 952 YRKKOMEKNIKPVDYKTP-----FNLKALPN 977
Db 1030 HRNAEKQHLEKENDYKQQLLIVTEDLRKTREDYEKELLRHADARSTLQKREDYTKALEQ 1089
QY 978 -EKVDGDRAAK-GHINIAETNNSVAVTPIRSEQQLHKSQSDVNLQPTSSKN-NF 1028
Db 1090 VEDLNKEIAALKAGINESQPFPISEKEDPLRQEVTVILKKQNAMLIQDQSSNLF 1143

QY 1039 CLLFLV 1044
 QY ::|:
 Db 1059 LKTYIV 1064

RESULT 9

A90551 conserved hypothetical protein MYPU_3130 [imported] - *Mycoplasma pulmonis* (strain UAB CT

C;Species: *Mycoplasma pulmonis*
 C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C;Accession: A90551
 R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001

A;Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulm*
 A;Reference number: A99512; MUID:21267165; PMID:11353084

A;Accession: A90551
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2819 <KUR>
 A;Experimental source: strain UAB CTIP
 C;Genetics:
 A;Gene: MYPU_3130
 A;Genetic code: SGC3

Query Match 4.0%; Score 212.5; DB 2; Length 2819;
 Best Local Similarity 18.2%; Pred. No. 0.12;
 Matches 246; Conservative 194; Mismatches 444; Indels 467; Gaps 56;

QY 1 MTKKHLKTLALA---LTTVSVTVYQEVYGLERBESV---KQQTQASASEDDWFEED 51
 3 MRNKILIGLGLASWAGIITTSAILPK--IGKGKDVAVWDNKNKQDSETSSGSVSDDFDKN 60

QY 52 NERKTNVKENSTVDETVSDLFSDGNSSNNSKTESVWSDPKQVPAKEVTQEAQNSN 111
 61 TKHSPDDKQKANQ---NKKKSSQSRSSESQ 96

QY 112 DASKVEVPKQD TASKETLETSTWEAKD--FVTRGDTL-VGF---SKS 153
 97 NFDSGFIDNNSTLSNOKRL---IKDKIRFVTFGDSIAAGFNAKMDYVPGYEVDPTK 150

QY 154 GINKLSQTSIHLVLP SHAADGTOI TOVASFAFTPDKRTAIAEYTSRLGE---NGKPSR 207
 151 QITGLSYSSYIADYIN DLPN KLESFKN FAFS--GTTLKDWNDYFNNKDTVFFNSTKSE 207

QY 208 LDIDQKEI I DGEI FFAVYQOLT KITI PINGYKSI GQDAFVDNKNIAEVNLPESLETISDYAF 267
 Db 208 -DLSQREDP-----IKRLT-----NSNLLNLIGAN-----DF 234

QY 268 AHMSLKQVKL PDNLKVIGELA FFDNQIGGKLYLP---RHLI---KLAERAFKSNRI 317
 Db 235 MSLAF EKFKEK DQVLSLIMKKDAS TNDLIG-VFILPILSSIRHEM GIRYKELISHIRKYNK 293

QY 318 QTVEFLGSKL-----KVIGEASFQDN-----NLRNVMLPDG LKIESEAFTG 359
 Db 294 VTI N LIGY PMPA LKLFNMVN KKLIGE SVKIGND TLLGFLIN FINSK I KDOV DLYENVNFVN 353

QY 360 NRGD-----EHVN-----NQVVLRT RTGQNP HOLATE-NTYVN PDK--SLWRA T- 400
 Db 354 AYDDPLWLKAKRNDFTDVA FDIHNE IGYKMAQE IFLKL AIDISK YNNP SEIN SSN ANY 413

QY 401 PDM DYT KMLE-EDFT YQNS-----VTGFSNK GLOK YTRRN KOLEIPKQHNGIT I TEGDN 454
 Db 414 1SYD H GKF SRL FEF K NKD SEIT TTVLGVN NSALFS KND KRIAL YES KF NT D NYSE I-- 470

QY 455 AFRN VDFO S KTLR KVD LEEIKL P S TIRK IGA FAF QOSN LKS FEA S DLEE I KEG A FMN R 514
 Db 471 LFR YSD I-FK T LSK S AVE-----FIN SS TFK TLD PEE DL--K KFL FANNE 513

QY 515 IGTLDL KDKL I KIGDA FPHIN H YIAV L PESV QEIG R AFR QNGA LHM FIGN KVK TIGE 574
 Db 514 ANY QKL FEL I LTKS R FV HNTI NRS IRDT L TLD NNG QAGI QIE LSH I MTF KK---E 568

QY 575 MAF LSNKL-----ESVNLS-----EOKQ KTI EVOAF S D NALS 607
 Db 569 ISNLD NYLGLLREF SSSSFYVDPETNLSFKDI KKLKNFLQKSEL KTTI LISKFTGAFIQ 628

QY 608 ETVLPP NTQ TIRE EAFK RNLKEVGS S TLSQIT FNA FQD QNDG DKR FGK KVVR-----661
 Db 629 NK ILS HNL SEL-----V L RIVS GEK T FALI-----EKL D DIF DN KAV YQ KVE KIE 674

QY 662 -----T HNN SHMLADGER FILD-----PD-----680

Db 675 DFL FAF VNT NEKA ITDF VRF VLNE I VAK P ENF KALV SFG I S QFLN SNEV I STQ TSQ AFY 734

QY 681 KLS S T MVD-----LEK VLK-----IIE-----G 698

Db 735 EVSK T L VNI VDTE B ELL EK VLK N F E I E I K S A Q K L S N P K E E N L I E K L Q K S I I S T F G F T N D 794

QY 699 LD YSTL ROTT-----O F R E M T A G K A L L S K S N L R Q E K O K F L Q B A O F F I G R V D L D K 751

Db 795 LG W S P E L K Q V V K S K S L S S V E O R D K F T S I L K E V V S F V E K S E V N K N K L I S A G T V I S N I E L F 854

QY 752 A I A K A E K A L V T K --K A T K N G H L L E R S I N-----K A V L A Y N-----N S A I K K A 791

Db 855 E N S D V F K G I F S H T I N F E K T P E L L N K I I D Y V I N S W E I D Q Y N D A F E F V L N F L K T N S S W F K T 914

QY 792 N V K R L E K E L D L I D L V E G K G P L A Q A T M V Q G V Y L L K T P L P L I P E Y Y I G L N V Y F D K S G K L I -- 849

Db 915 Y V K -----D I F S S V N S E I Q Y -----N I L T S K L N K L A S Q 944

QY 850 Y A L D M S D T I G E Q O K D A V G N P I L N V D E D N E G Y H T L A V A T I L A D Y E G L Y I K D I L I N S S L D K -- 906

Db 945 A G F E L T N S S V E S I K N I F T I N T L K I A D S N -----T I D A L V D --T F V N S K T L I E I D K K N F 995

QY 907 --I K A I R Q I P L A K Y H R L G I F O A T R N A A E A D R -----936

Db 996 I A S L K N I Q I P A L S D D N F F S V F K A L I S S N K L K D K A F V N Q Y K Q D I T I N V L K E L F E E K F F N L I 1055

QY 937 --L I P K T P K G Y -----I N E V P N Y R K K O M E K N L K P -----963

Db 1056 Y S F M M P K T I S E L F P E N N P E A L N K T T I I R N A L K K N L D P D F I N K M M D I I F D N I L D K Y Q K L E N Y G 1115

QY 964 -----V D Y K T P I F E N K A L P N E K -----V D G D R A A K G H N I N A E T N N S V A 1000

Db 1116 Q V I G L I L K D N S D A I A T K V Q A I F K D V V N E A N F Q D V L K V L L V S Q T K K F M G F E L N A D E T N K I S 1175

QY 1001 V T P P I R S E Q O L H K S Q S D V N L P Q T S S K N N F I Y E 1031

Db 1176 T D L T K I L Q V F E S N N F A K S L A L A T K D S I V D 1206

RESULT 10

T39068 Query Match 3.9%; Score 209; DB 2; Length 1208;
 Best Local Similarity 19.4%; Pred. No. 0.051;
 Matches 220; Conservative 170; Mismatches 367; Indels 378; Gaps 52;

C;Species: *Schizosaccharomyces pombe*
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T39068
 R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Connor, R.E.; Wood, V.
 submitted to the EMBL Data Library, August 1997

A;Reference number: Z21825

A;Accession: T39068
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-1208 <MUR>
 A;Cross-references: EMBL:Z81317; PIDN: CAB03608.1; GSPDB: GN00066; SPDB: SPAC6G9.06C
 A;Experimental source: strain 972h-; cosmid c6G9
 C;Genetics:
 A;Gene: SPDB: SPAC6G9.06C
 A;Map position: 1

QY 20 TYSQEVIGLEREESVKOEQTOSA-----SEDDWFEEDNERKTNVSKENSTVDETVSDLFS 74
Db 93 SFDPNSYGL---SAISKOATOEALISISQNDSY---DVSKLTDISK-NSEIDHTDGEI-- 143
QY 75 DGNSNNSSSKTESVSD-----PKQVPA---KPEVTQEASN 108
Db 144 PANAALTIREQEKVLEKVSRENFGIRIKIVCLEKRLESMAPEQIKEAVKDNVELHAERAN 203
QY 109 SSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGESKSGINKLSQ-----T 161
Db 204 LQLQLKRTESSLQKSEDKNFKLE---EKVDYLSK-----VNDVEQSQNWKVFT 248
QY 162 SHLVLPISHAADGTQLTQVA-SRAFPDQKTAIAEY-----TSRIGENGKPSRLDIDQK 213
Db 249 ERIRFLENALEKVOREKDSLSTEMEEDKSNKEVDYEEYETRQLQNLDELSE--ELDVAQ- 305
QY 214 ETIDEGETFNAYQLTQLTIPNGYKSIGQDAFVDNKNIAEVNLPE-----PKQVPA--- 257
Db 306 DLLTEKE----DEIATLKRQIEEKENSSAFENEENSSYVHLOQDYAILQACDEFADRI 361
QY 258 -----SLETISDYFAFAHMSLKVQVLPDNLKVGELAFFDNQIGGKLYLPRHLIKLAERAF 312
Db 362 QVLTADILEKEKENQIMHESEASIGLTDQV-----H-TLQEQLH 400
QY 313 KSNRIQTVFGLGSKLKVIGE--ASFQDNNLRNVMLPDLGKIESEAFTGNPGDEHYNNQV 370
Db 401 KAN--EEEFLHDQISRMNEEGKNFEDIMLQFRSLLEERDVLESKLOT-----LEDDNNSL 454
QY 431 VLRTRT-GQNPQLATENTYVNPDKSLWRATPDMDYTKMLEEDFTYQKNSVTGFSNKGLO 429
Db 455 RLMTSSLGNQIESLRTONREIDEEKHNLRLLASKNSDKALAE-----TNIRLQ 502
QY 430 KVRRNKNLEIPKOHNGITITEIGDNAFRN-----VDFQSKTLRYDLE 472
Db 503 EV--TKELETLRMKNSNDLNEIHDLRENEEGLTLKIDSITKEKDRLINEQRIKSYEVN 560
QY 473 EIKLPSTIRK-----IGAFAFQSNNLKSFEASEDLEEIKEGAFMNRRIGT 517
Db 561 VSELNGTIDEYRNKLKDKEETYNEWMAFQYKDNDLRRFHESINKLQDREKELTSN--- 616
QY 518 LDLKDKLIKIGDAAFHINHIYATIVLPESVQEIGRSAFRONGALHIMFIGNKVKTIGEMAF 577
Db 617 LEKKNLVI-----SSLRETVAMLEKERESIKKLYLSGN----- 648
QY 578 LSNKLESVNLSE-----OKQLKTI--EVQAFSDNALSEEVVLPPNLQOTIREEAFKRN 626
Db 649 -AKDLDNTNLMEIILNDKISVVLQRQLTDVKDELDVSEEEEREATVAGQKULSASFELMSNEK 707
QY 627 HILKEVKGSSTLISOITFNAFDONDGDKR-----FGKKV-----VRTHNNSHMLADGERFI 676
Db 708 QALELKYSSSLKNEI-INA--QNLIDRREEELSESKKLFEERKIRSGSNDDIEKNKEINV 764
QY 677 ID---PDKLS-----STMVDLEKV-----LKIIEGLDYSTLR- 705
Db 765 INSELADKLAQIRHLESDDKMELDKLVHHLNRGIEEANIEENAVKRLCLLMGCDYSSVSI 824
QY 706 -----QTTQTOFREM-----TAGKALISK 725
Db 825 LQTVSQLEHFVNQQTIRSLKQELRHDFVQFSGKKEQELSRSFEKFGLGTTKHDILAQ 884
QY 726 SNLROGEKQKFLQE-AQFFFLGRVD-----LDKAIKAEKALVTKATK 767
Db 885 RNRNVSEKMDLENAQKFSSPDRKNGYLYPSEHTSKIELEKTIEDLKL-----QDELK 942
QY 768 NGHLLERSINKAVLAYNNSAIK-KANVKRLEKELDDLLTD-----LVEGKG 811
Db 943 NRNLMLMDDIS---SYNKQTTKLOEKIKWLERERSILIDELESYRSNQFNQNYQNNLVQDKN 998
QY 812 PLAQATMVQGVYILKTPPLPEYYIGLNVY---FDKSGKLIYA--LDMMSDTIGEGQKDAY 866
Db 999 ELEER-----LKEIQKELEVYNNHFMQAEALMTSNTDESQMLKTLREAL 1044
QY 867 GNPIILNTDEDNEGTYHTLAVATIADYEGLYIKDILNSSLKIKAIROIPLAKYHRL 921

QY	381 HQLATENTTYVNPDKSLWRAFPDMYTKWLEDFPTYQKNSVTGFSNKGLOK--VRRNKNLE 438	Query Match	3.9%; Score 209; DB 2; Length 6713;
Db	462 KLVCECNDLQTLTQ-----RIDLRCQIQ--YLLITNSVSNSKGPIRKKEEIQFIQNT- 512	Best Local Similarity	17.6%; Pred. No. 0.65;
QY	439 IPKOHNGTITTEIGDNAF--RNVDFOSKTLRKYDLEIKLPISTIRKIGAFAFQSNLKS 495	Matches	199; Conservative 207; Mismatches 406; Indels 318; Gaps 46;
Db	513 -MOEDDTITESDSQKUTERLIVEF--KNIQLOERNAELLKVRNL-ADKLESKEKS 567		
QY	496 FEA-----SEDEEIKEGAFMNRRIGTLIDDKLKLIGDAAFHINHIVAIVLPESSVQETG 550		
Db	568 KQSLQKIESETVNEAKE-AIITIJKSEKMDLIESRIE-----LQKELHL- 610		
QY	551 RSAFRQNGAHLFIGNKVKTIGEMAFLSNKLESVNLSEQKQKLTIEVQAFSDNALSEWV 610		
Db	611 -----KTSVPNEDASYN-VTIKQLTETKRDLESQVO- 641		
QY	611 LPPNLQQTIREEAFKRNHLKEVKGSSTLSQITFNAFDQNDGDKRFGKKVVRTHNNHMLA 670		
Db	642 ---DLQT-----RISQITRESTENMSLINKETIQDLYSDSKSDISIKLGKEKSSRILA 689		
QY	671 DGERFLIDDPDKLSSTMVDELEKVLKLTIEGLDYLSTLRQTQO--FREMTTAGKALLS--- 724		
Db	690 E-ERFKLISNTLDLTKAENDQQLRKRFDYLQNTILKQDKTHETLNNEVSC-KSKLIVET 747		
QY	725 -----KSNLR-----QGEKOKFLOEAQFFLR 746		
Db	748 ELLNLKEEQKLRLVHLTEKLNKQELNKLSPEDSLRIMVTQLOLQEREDLLET----R 802		
QY	747 VDLDKAIAKAEKAL-VTKKAT--KNGHL--LERSINKAVLAYNNSAIKKANVKRLEKLD 801		
Db	803 KSCQKKIDLEDALESLKKETSQKDHHKQLEEDNNNSIEWWQN-----KIEALKKDY 856		
QY	802 LLTDLVEGKGPLAQATMVGQVYLLKTPPLPLPEYYIGLNVYFDKSGKLIYALDMSDTIGE 861		
Db	857 SVITSVDSK---QTDIEKLQYKVKS-----LEK----- 881		
QY	862 QKDAYGNPILNVDEDNEYHTLAV--ATLADYEGLYTIKDILNSSLDKIK---AIRQIP 914		
Db	882 -----EIEEDKIRLHTNVMDETIND-----DSRKELEJKSKINTLTDAYSQIK 924		
QY	915 LAKYHRUGIFQAIRNAAEADRLLPKTPKGYLNEVPNVRKKOMEKNLKPVDYKTPFLN- 972		
Db	925 EYK---DLVETTSQSLQQTINSKULDESFKDFTNQINKLTDEXTSLEDKISLILKEQMFNL 980		
QY	973 -----KALPNEKVGDRAAKGHNINAETNSVAVTIRS--EQQLHKSQSDVN--LP 1020		
Db	981 NELDLQKKGMEKEKADPKK---RISILOQNNEKEVAVKSEYESKLISKIQNDLDDQQTYY 1035		
QY	1021 QTSSKNNFIYEI 1032		
Db	1036 ANTAQNNIEQEL 1047		
RESULT 12			
B89921	hypothetical protein ebha [imported] - <i>Staphylococcus aureus</i> (strain N315)		
C;Species	<i>Staphylococcus aureus</i>		
C;Date	10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001		
C;Accession	B89921		
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguda, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, H.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.			
Lancet	357, 1225-1240, 2001		
A;Title	Whole genome sequencing of methicillin-resistant <i>Staphylococcus aureus</i> .		
A;Reference number	A89758; MUID:21311952; PMID:11418146		
A;Accession	B89921		
A;Status	preliminary		
A;Molecule type	DNA		
A;Residues	1-6713 <XJR>		
A;Cross-references	GB:BA000018; PID:913701232; PIDN:BAB42527.1; GSPDB:GN00149		
C;Experimental source	strain N315		
C;Genetics			
A;Gene	ebha		
QY	29 ERTEESVKQEQTOSASEDDWFEEDNERKTNVSKENSTVDETVSDLFSDGNSNNSSKTE-- 86		
Db	5190 DKKEAVDQALQAAQSITDPTNGSNANKDAVEQALTQKQEVKVNEL--NGNERVAEAKTOAK 5247		
QY	87 -----SVVSDPKQVKPKAkp--EVTOEASNSNDASKVEVPQDTASKET 129		
Db	5248 QTIDQLTHLNADQIATAKQNDQATKLQPIAELVQDQATQLNQSMQDQLQQAVNEHANVEQT 5307		
Db	5308 IDYTOADSQKAVKOATAADAENV---KQANKQOVDQALQNL-----AK 5352		
QY	130 LETSTWEA-----KDFVTRGDTLVGFSKSGINK--LSQTSHLVLPSPHAADGTQLTQVAS 181		
Db	5538 QALNGDERRVALAK-----TNGKH--DIDOLNALNNAQDGFGRIDOSNDLNQIQQIV 5403		
Db	5404 DEAKALNRAMDOLSQERTGNEGRTRKGSTNTYVNADTQVKQVYDEAVDKAKQALDKSSGQNL 5463		
QY	271 SLKQV-KLPDNL---KVIGELAFFDNQIGGGKYLPRHLIKL-AERAFKSNRIFTQVEFL 323		
Db	5464 TAEQVIKLNDAVTAAKKALNGEERLNNRKAELQRLDQLTHLNNAQROLAIQQINNAETL 5523		
QY	324 GSKLKVIGEASFQDNLNRLNVMLPDGLEKIESEFTGNGPGEDEHYNNQVLRTRGQNPQHL 383		
Db	5524 NKASRAINRATKLNDAM-----GAVQQYDDEOHILGVISTSTNYIN----- 5562		
QY	384 ATENTIVNPDKSLWRAFPDM-----YTKWLEDFPTYQKNSVTGFSNKGLOKRR 433		
Db	5563 ADDNLKANYDNATAANAAHELKVQGNAIAKAAEQLKQNTIDAQNALNGDQNLANAKDA 5622		
QY	434 N-----KNELEIPKQHNGITITBEGDNARFNVDFOSKTLRKYDLEEIKLPISTI 480		
Db	5623 NAUVNSLNGLNLQQQDLAHKAINNADTVSDVTIDVNNQIDNDAMETLKHLDNEIPNAE 5682		
QY	481 RKIGAFAFQSNLKSFEASEDLEETIKEGAFMNRRIGTLIDDKLKLIGDAAFHINHIVAI 540		
Db	5683 QTVNYQNADDNAKTNF---DDAKRILANTLLNSDNTNVNDINGAIQAWNDATHNLDQRL 5739		
QY	541 --VLPESSVQEEIGRSAFRQNGAHLFIGNKVKTIGEMAFLSNKLESVNLSEQKQL---K 594		
Db	5740 QDAKDKAQSIQNA-----LANKLK-----EIEASNATDQDKLIAKNK 5777		
QY	595 TIEV-----QAFSDNALSEWVLPNLTQITREEAFKRNHLKEVKGSSTLSQITENAF 645		
Db	5778 AEELANSIINNINKATSNQAVSQV-----QTAGNHAIEQVHANEIPKAKI----- 5822		
QY	646 DQNDGDKREGKVVVRTHNNHMLADGERFTIDDPKLSSTMVDLEK-VLKTIIEGLDYSTL 704		
Db	5823 ---DANKDVKQV-----QALIDDEIDRNPNTDKEQKALK-----DRI 5857		
QY	705 RQTTQTQFREMTTAGKALLSKSNLROGEKO--KFLOEAQFFL-GRVFLDKAIAKAEKALV 761		
Db	5858 NQILQQGHNDINNA---LTKEEEIQAKAQLAQALQDIDKLVAKEDAKQDVQVQALI 5913		
QY	762 TK-----KATKNGHLIERSINKAVLAYNNSAIKKANVKRLEKELDLTDLVECKGPLAQ 815		
Db	5914 DEIDQNPNTDKEKQALKDRINO-1LQOQHNGINNAMTE-----EIEQAKAQLAQ 5963		
QY	816 ATMVQGVYLLKTPPLPEYYIGLNVYFDKSGK-LTIALDMSDTIGEQQKDAYGNPILNVD 874		
Db	5964 A-----LKEIKDLVAKENAKQDVQVQALIDEIDQNPNTDKEKQALKDRINOQIL 6015		
QY	875 EDNEYHTL-----AVTADYEGLY-IKDLNSSLKIKATRQIPLAKYHRLG 922		
Db	6016 Q---QGHNDINNAMTKEEIQAKAQLA--QALQDIDKLVAKEDAKNAKALANAK---- 6066		

Qy	923	IFQAIRNAAEADRLLPKTPKGYLNEVNPYRKQMEKNUKPVDPYKTPFENKALPNEKVDG	982	Qy	544	E---SVQEI-----GRSAFR--ONGALHLMFIGNKVKTIGEM-----AF	577
Db	6067	-----RDOINSNPDLTPEOKAKALKEID-----EA	6091	Db	411	EVANTKEVKTLILDFYHAAAGQTTLQDYKNDPNAQVQDDVVATSLMLKTRDCKTLKA	470
Qy	983	DRAAKGHNINAETNNSAVTPIRSEQOLHKSQSDVNLPQTSSKNNFIYEL	1032	Qy	578	LSNKLESV-----NLSEQKQLKIEVQAFSDNAISEVVLPNQLTIREEAFKR	625
Db	6092	EKRALQNVNAQTID-----QLNRG---LNLLDDIRNTHWNEV	6127	Db	471	LQDKTDSILNSLNINSGIGNIDDYSKQLEAVDASKLEAVND---DIKKIKADKGRD	525
RESULT 13							
C97033 uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylicum							
C;Species: Clostridium acetobutylicum C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001							
C;Accession: C97033 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum A;Reference number: A96900; MUID:21359325; PMID:21359325 A;Accession: C97033 A;Status: Preliminary A;Molecule type: DNA A;Residues: 1-1227 <KUR> A;Cross-references: GB:AE001437; PIDN:AAK79054.1; PID:915023995; GSPDB:GN00168 A;Experimental source: Clostridium acetobutylicum ATCC824 C;Genetics: A;Gene: CAC1080							
Query Match Best Local Similarity 18.9%; Pred. No. 0.055; Matches 221; Conservative 164; Mismatches 363; Indels 419; Gaps 54;							
Qy	3	KKHHLKTLALALITTVSVVTVIYQEVYGLEREEBSVKQEQTQSASE-DDWFEEDNERKTMUSKE	61	Qy	684	STMVDELEVKLIEGLDYSTLQTTQFREMTTAGKAL---LSKSNLR-----	729
Db	5	KNILLVFSISVTSAVLLSKPAKAYAADNSVYKSIINSNSNEINVKGKEVQEVNWNK	64	Db	582	GITITIENIKVIE----PIVOLSEVYVRIIVGVTWVUDYKTLGINVNNDNNIIVNAE	636
Qy	62	NSTVDETWSDFSDGNGNNSSSKTESVSDPKQVKPKAEPVTOEASNS-----NDAS	114	Qy	626	NHLKEVGS--STLSQITNAFDONDGDKRFGGKKVVRTHNSHMLADGERFIIIDPKLS	683
Db	65	N-----DSNSKVSSSENQVSNKENSNPKVSSSEIQSINKNVNLQVQDNKNS	112	Db	526	LTIQEIRDSVKKTIDYIN-STSNVSKGDGSVSDYITIGIDGVTEINIE--FVNERIKES	581
Qy	115	KVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSSKGINKLQTSHTHLVLPSHAADGT	174	Qy	778	KAVLAYNNSAIKKANVKRLEKELDLITLIVEGKPLAQATMVQGVYLLKTPPLPEYYIG	837
Db	113	VLAASNVDEVKK--INSGNVQTSYIAIGETKV-----	144	Db	685	QDILDYVNADLKLQY---TDLQNKYKTA-DEVIRVQAQETIYRALMQINLGKATTADY	735
Qy	175	QLTQVASFATPDKTKTAIAEYTSRIGENGKPSRLDIDQKEITID-----EGEIFNA	224	Qy	778	LNVYFDKSGKLIYALDMSDTIGEOKDAYGNPLNVEDNEGYH---TLAVATLADYE	892
Db	145	-----KPTDLDINKATVDARSAEGTDLSVEYEIYD	175	Db	736	LGLTDINDGLLHYA--TDLQNKYKTA-DEVIRVQAQETIYRALMQINLGKATTADY	792
Qy	225	YQLTKLTIPNGYKSIGQDAFVDNKNIAEVLPESLETISDYFAAHMSLKVQLPDLNKLV	284	Qy	893	GLYIKDILNSLDKIKAIROQIPLAKYHRLGIFQAIRNAAEADRLLPKTPKGYLN--EVP	950
Db	176	VSQTAET-----QEAFT-----RNINDGVANVSDYI-----JL	203	Db	793	TIGIIDVNDSILTYVNADLQ-----	822
Qy	285	GELAFFDNQIGGKLYPRHLIKLAERAFAKSNRIQTVEFLGSKLKVIGEASFQDNNLRNVM	344	Qy	951	NYRKOMEKLNKPVDPYKTPIFNKALPNEKVDGDRAA---KGHNINAETNNSAVTPIRS	1006
Db	204	G---ATFVNDAN-----LDSVNKFYHKRYATV---TKFKDV--ATKTSNALKNI-	245	Db	823	--KAEIEKNIQ-----IYNALI---KIDSGSATIDDYRTIGITTTVIDNISYVNIRI	869
Qy	345	LPDGLEKIESEAFTGNGPDEHYNNQVVRTRTGQNPQHOLATENTYVNPDKSLWRATPDMD	404	Qy	1007	EQQLHKSQSDVNLPQTSSKNNFIYEL	1033
Db	246	-----NN-----GQG-----GETD	254	Db	870	K-----GMNIPQVSDAKRFINIL	888
Qy	405	YTKWLEEDFTYQKNSVTGFSNKGLOQKVRRNKLEIPKQHNG--ITITEIGDN-----	454	Qy	544	E---SVQEI-----GRSAFR--ONGALHLMFIGNKVKTIGEM-----AF	577
Db	255	YTA-LE-----VSGVTQPYLDV--NKNIVKERQKGRDLTITEISDSAASIARI	302	Db	411	EVANTKEVKTLILDFYHAAAGQTTLQDYKNDPNAQVQDDVVATSLMLKTRDCKTLKA	470
Qy	455	--AFRNVDFOSKTIRKY-----DLEBIKL-----PSTIRKIGAFAFQSN	491	Qy	578	LSNKLESV-----NLSEQKQLKIEVQAFSDNAISEVVLPNQLTIREEAFKR	625
Db	303	NTALDNMDAGVATLVEDYQATIGANNVQPLHVADEVNLSAMDQRWGDVSEADGINTIMTYIN	362	Db	471	LQDKTDSILNSLNINSGIGNIDDYSKQLEAVDASKLEAVND---DIKKIKADKGRD	525
Qy	492	NLKSFEASEDLEEIKEGAFMNMRIGLD-----IKDKLIKIGDAAFHINHIVIAVLP	543	Qy	626	NHLKEVGS--STLSQITNAFDONDGDKRFGGKKVVRTHNSHMLADGERFIIIDPKLS	683
Db	363	NINSGVGTED-DYINSHAVDSNE-GNIDYDILNANIIEKKTAKGQD-----LTIP	410	Db	526	LTIQEIRDSVKKTIDYIN-STSNVSKGDGSVSDYITIGIDGVTEINIE--FVNERIKES	581

Query Match	3.9%	Score 207.5; DB 2; Length 1119;	Qy	544	E---SVQEI-----GRSAFR--ONGALHLMFIGNKVKTIGEM-----AF	577
Best Local Similarity 19.4%; Pred. No. 0.053; Mismatches 399; Indels 341; Gaps 59;	Matches 223; Conservative 184; Mismatches 399; Indels 341; Gaps 59;	Qy	411	EVANTKEVKTLILDFYHAAAGQTTLQDYKNDPNAQVQDDVVATSLMLKTRDCKTLKA	470	
		Qy	578	LSNKLESV-----NLSEQKQLKIEVQAFSDNAISEVVLPNQLTIREEAFKR	625	
		Db	471	LQDKTDSILNSLNINSGIGNIDDYSKQLEAVDASKLEAVND---DIKKIKADKGRD	525	
		Qy	626	NHLKEVGS--STLSQITNAFDONDGDKRFGGKKVVRTHNSHMLADGERFIIIDPKLS	683	
		Db	526	LTIQEIRDSVKKTIDYIN-STSNVSKGDGSVSDYITIGIDGVTEINIE--FVNERIKES	581	
		Qy	684	STMVDELEVKLIEGLDYSTLQTTQFREMTTAGKAL---LSKSNLR-----	729	
		Db	582	GITITIENIKVIE----PIVOLSEVYVRIIVGVTWVUDYKTLGINVNNDNNIIVNAE	636	
		Qy	730	-QGEKOKFLOEAQF---FLGRVDDLKATAKEAKL---VTKKATKNGHLERSIN	777	
		Db	637	LNKDKDVKIQDQTRVDNTINNIDVINKIGAGDAVLSDYFVNIGIT---DVY	684	
		Qy	778	KAVLAYNNSAIKKANVKRLEKELDLITLIVEGKPLAQATMVQGVYLLKTPPLPEYYIG	837	
		Db	685	QDILDYVNADLKLQY---TDLQNKYKTA-DEVIRVQAQETIYRALMQINLGKATTADY	735	
		Qy	838	LNVYFDKSGKLIYALDMSDTIGEOKDAYGNPLNVEDNEGYH---TLAVATLADYE	892	
		Db	736	LGLTDINDGLLHYA--TDLQNKYKTA-DEVIRVQAQETIYRALMQINLGKATTADY	792	
		Qy	893	GLYIKDILNSLDKIKAIROQIPLAKYHRLGIFQAIRNAAEADRLLPKTPKGYLN--EVP	950	
		Db	793	TIGIIDVNDSILTYVNADLQ-----	822	
		Qy	951	NYRKOMEKLNKPVDPYKTPIFNKALPNEKVDGDRAA---KGHNINAETNNSAVTPIRS	1006	
		Db	823	--KAEIEKNIQ-----IYNALI---KIDSGSATIDDYRTIGITTTVIDNISYVNIRI	869	
		Qy	1007	EQQLHKSQSDVNLPQTSSKNNFIYEL	1033	
		Db	870	K-----GMNIPQVSDAKRFINIL	888	
RESULT 14			Qy	544	E---SVQEI-----GRSAFR--ONGALHLMFIGNKVKTIGEM-----AF	577
			Db	411	EVANTKEVKTLILDFYHAAAGQTTLQDYKNDPNAQVQDDVVATSLMLKTRDCKTLKA	470
			Qy	578	LSNKLESV-----NLSEQKQLKIEVQAFSDNAISEVVLPNQLTIREEAFKR	625
			Db	471	LQDKTDSILNSLNINSGIGNIDDYSKQLEAVDASKLEAVND---DIKKIKADKGRD	525
			Qy	626	NHLKEVGS--STLSQITNAFDONDGDKRFGGKKVVRTHNSHMLADGERFIIIDPKLS	683
			Db	526	LTIQEIRDSVKKTIDYIN-STSNVSKGDGSVSDYITIGIDGVTEINIE--FVNERIKES	581
			Qy	684	STMVDELEVKLIEGLDYSTLQTTQFREMTTAGKAL---LSKSNLR-----	729
			Db	582	GITITIENIKVIE----PIVOLSEVYVRIIVGVTWVUDYKTLGINVNNDNNIIVNAE	636
			Qy	730	-QGEKOKFLOEAQF---FLGRVDDLKATAKEAKL---VTKKATKNGHLERSIN	777
			Db	637	LNKDKDVKIQDQTRVDNTINNIDVINKIGAGDAVLSDYFVNIGIT---DVY	684
			Qy	778	KAVLAYNNSAIKKANVKRLEKELDLITLIVEGKPLAQATMVQGVYLLKTPPLPEYYIG	837
			Db	685	QDILDYVNADLKLQY---TDLQNKYKTA-DEVIRVQAQETIYRALMQINLGKATTADY	735
			Qy	838	LNVYFDKSGKLIYALDMSDTIGEOKDAYGNPLNVEDNEGYH---TLAVATLADYE	892
			Db	736	LGLTDINDGLLHYA--TDLQNKYKTA-DEVIRVQAQETIYRALMQINLGKATTADY	792
			Qy	893	GLYIKDILNSLDKIKAIROQIPLAKYHRLGIFQAIRNAAEADRLLPKTPKGYLN--EVP	950
			Db	793	TIGIIDVNDSILTYVNADLQ-----	822
			Qy	951	NYRKOMEKLNKPVDPYKTPIFNKALPNEKVDGDRAA---KGHNINAETNNSAVTPIRS	1006
			Db	823	--KAEIEKNIQ-----IYNALI---KIDSGSATIDDYRTIGITTTVIDNISYVNIRI	869
			Qy	1007	EQQLHKSQSDVNLPQTSSKNNFIYEL	1033
			Db	870	K-----GMNIPQVSDAKRFINIL	888
RESULT 14			Qy	544	E---SVQEI-----GRSAFR--ONGALHLMFIGNKVKTIGEM-----AF	577
			Db	411	EVANTKEVKTLILDFYHAAAGQTTLQDYKNDPNAQVQDDVVATSLMLKTRDCKTLKA	470
			Qy	578	LSNKLESV-----NLSEQKQLKIEVQAFSDNAISEVVLPNQLTIREEAFKR	625
			Db	471	LQDKTDSILNSLNINSGIGNIDDYSKQLEAVDASKLEAVND---DIKKIKADKGRD	525
			Qy	626	NHLKEVGS--STLSQITNAFDONDGDKRFGGKKVVRTHNSHMLADGERFIIIDPKLS	683
			Db	526	LTIQEIRDSVKKTIDYIN-STSNVSKGDGSVSDYITIGIDGVTEINIE--FVNERIKES	581
			Qy	684	STMVDELEVKLIEGLDYSTLQTTQFREMTTAGKAL---LSKSNLR-----	729
			Db	582	GITITIENIKVIE----PIVOLSEVYVRIIVGVTWVUDYKTLGINVNNDNNIIVNAE	636
</td						

QY 35 KOEQTOSASED-DWFEEDNERKTKNVSKENSTV-----DETVDLFDGNS----- 78
 Db 94 KDLORDSANQILDKIKGQDNTKTNV-ENFDIAFNRYIKDSTITENSDRNDDVGIRDE 152
 QY 79 -----NNSSSKTRSVVSDPK-----QVPAKPEV--OEASN 108
 Db 153 DISEFKKSKEPEKPNTPKEEDQIIOSPNPKLSVNDQKNLFLNIEKLKKNLGSKSEN 212
 QY 109 SSNDASKVVPKQDTASKKETLETSTWEAKD---FVTRGDTLVGFSKSGINKLS---QT 161
 Db 213 ILNDSOKIENDKONTNLSEKEKNSENILKTPDNNPSKSYNNNN-----: 266
 QY 162 SHLVLPSHAADG----TQLTOVASFAPTDPDKTIAEYTSRSLGENGKPSRLIDQKRI 215
 Db 267 SELSPPSOTIGKIVPRYPSYLIKELYRIDDINTG---RVTLGKN----RL---KEL 314
 QY 216 IDEGEIFNAYQ-----LTKLTIPTYKSIGQDAFVDNKNAEVN- 254
 Db 315 IKKG-LSNKEQVNELIENSKEASNLILTLIKKDEPNLINIPKDPY--KELIFQDLK 371
 QY 255 --LPESLLETISDYFAHMSLKVQLPDN----LKVIIGELAFFDNOIGGKLYLPRHL 305
 Db 372 EDKKPQYLEDIKSKVH--SICKPIDLENTKSRSQAIKDLNE--FLKN----PND-A 418
 QY 306 KLAERAFKSNRIOTVEFLGSKLKVIGEASFFODNNLRNVMLPDGLKTESEAFTGNPGDEH 365
 Db 419 QASKTLAQANKI-QHLEDLKSKVHSIKPIDLENTKSRSQAIKDLN----EP-LKNNPN 524
 QY 366 YNNQVVLRLTRTGPOLATENTTYVP----DKSLWRAATPDMDYTKMLEEDFTYQKNSV 420
 Db 474 ASKTLAQANKI-QHLEDLKSKVHSIKPIDLENTKSRSQAIKDLN----EP-LKNNPN 524
 QY 421 TGFSNKGLQKVRRNKLTEPKOHNGITITEIGDNAFRNVDFOSK--TRLKYDLEEIK-LP 477
 Db 525 DAQASKTLAQANKIQLHE-----DLKSKVHSIKPIDLENTKSRSQ 563
 QY 478 STIRKIGAFAFQSNNLKSFEASEDLEEIKEGAFMNR-IGTL--DLKDKLJIKIGDAFHINH 536
 Db 564 QAIKDLNEFL--KNNPNDQASKTLAQ-----NKTQHLEDLKS-----VHS 604
 QY 537 IYATIVL-----PESVQEIGRSAFRQNG----ALHLMFIGNKVKTIGEMAFLSNKLEVY 585
 Db 605 IKPIDLENTKSRSQAIKDL--NEFLKNNPNDQASKTLAQANKIQLHEDLKSKVHSIKPI 662
 QY 586 NLSEQKQQLKTEVQAFSDNALSREVVL--PPNLOTIR--EEAFKRMHLKEVKG----- 633
 Db 663 DLENTKSRSQ----QAIKD--LNEFLKNNPNDQASKTLAQANKIQLHEDLKSKVHSIKPI 716
 QY 634 --SSTLSQ----ITFNAFDQNDGDKRGKVVVRTHNNHMLADGERFILIDPDKLSSTMV 687
 Db 717 DLENTKSRSQATKDNFELKNPNDQASKTLAQAYENNGDLKAKENAYEKITIKLTNTQE 776
 QY 688 D-----LEKVULKIEGLDYSLTROTQTOFREMTAGKALL--LSKSNLROGEKOK 735
 Db 777 DHYKLGIIRFELKYYEHSIESP-OTIKLDPKHK-KALHNKGIAAMMLNNKNAESFEK 834
 QY 736 FLOEAQFFLGRVLDKAIAKAEXALVTKK-----ATKNGHILLERSINKAV---- 780
 Db 835 AIQ-----IDKNYGTAYQKGIAEKKNGDMQQAFASFKNAYNLKDPNPNYALKAGIVS 886
 QY 781 -----LAYNNSAIKKANVRLEKEDDLTDVE----- 808
 Db 887 NNLGNFKQSEEYLNFFNANAKKPNEIATIYNLSIAKFENNKLEESJETINKAIDLNPEKSE 946
 QY 809 -----GKGPLAQATMVQGVLLKTPPLPEYYIGINVYFDKSGKLIYALD-MSDT 857
 Db 947 YLYIKASINLKKENYQNAISLYSLVIEKNP-ENTSAVNLAKAYEXSGNSQASISTLEK 1005
 QY 858 IGEQOKDAYGN-----PILNVEDDENEYHTLAVATLADYEGLYIK 897
 Db 1006 INKNKNKLAINNLGILYKKEKNYQKAIEIFEKAIIN-SDIEAKYNTLATTIEINDNTRAK 1063

QY 898 DILNSSLDKKIAIRQIPLAKYHRIGFOAIRNAAEADRLLPKTPKGYLNEVPNRYKKQM 957
 Db 1064 DLIR-EYTKLKPNNPEAL---HALGIIENNNNDQTLREL-----IKKFPNYKKNE- 1111
 QY 958 EKNLKPV 964
 Db 1112 --NIKKI 1116

RESULT 15
 T28677
 Rhopty protein - Plasmodium yoelii
 C;Species: Plasmodium yoelii
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
 C;Accession: T28677; C45521
 R;Keen, J.; Sinha, K.; Brown, K.; Holder, A.
 Mol. Biochem. Parasitol. 65, 171-177, 1994
 A;Title: A gene coding for a high molecular mass rhopty protein of Plasmodium yoelii.
 A;Reference number: Z20508; MUID:95021522; PMID:7935623
 A;Accession: T28677
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-2269 <KEE>
 A;Cross-references: EMBL:127838; NID:9457145; PID:9457146; PIDN:AAA21304.1
 R;Keen, J.; Holder, A.; PlayFair, J.; Lockyer, M.; Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A;Title: Identification of the gene for a Plasmodium yoelii rhopty protein. Multiple
 A;Reference number: A45521; MUID:91101660; PMID:2270106
 A;Accession: C45521
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 2131-2269 <KEE>
 A;Cross-references: GB: M34283

Query Match 3.9%; Score 207; DB 2; Length 2269;
 Best Local Similarity 18.8%; Pred. No. 0.16; Mismatches 226; Conservative 206; MisMatches 432; Indels 340; Gaps 58;
 Matches 989 DHDYNIKELKSHDKSNGYKTEADQNKKAQKNNKLFEQYKEEVTLVNLKYYAVELKNF 1048
 QY 29 EREESVQEOQTOSASEDDWFEEDNERKTKNVSKENSTVD--ETVSDLFDGNSNNSSSKT 85
 Db 86 ESVVSDPKQVPK----AKPEVTOEASNNSNDASKV-----EVPKQDTASK---- 127
 QY 1049 DKTKNDSKQIKEIKDAHNYCTLESGKSEKKMNEIKNEKIHIEDEVANNDKSNAIKTSIK 1108
 Db 128 --ETLETSTWEAKDFTVTRGDTLVGFSKSGINKLSQTSHLVPLSHAADGTQLTQVASFAF 184
 QY 1109 VSVEPFKTKIJKINEIRTKSD-----DCIKETNDLE-----KOISNL-- 1145
 Db 185 TPDKKTAIAEYTSRSLGENGKPSRLIDQKEIIDEGEIFNAYQLTKLTIPTYKSIGQDAF 244
 Db 1146 -----SIDTQETKLTENGK-----OLKTLE--BILLESLKKQKKNED----- 1180
 QY 245 VDNKNTIAEVNLPESELETISDYFAHMSLKVQLPDKNLKVIGELAFFDNOIGGKLYLPRHL 304
 Db 1181 -OKKELDEVN--SKINNIENTVNQHKKNYEIGIVEKINETAKTN--KNOIESTKELIKPT 1235
 QY 305 IKLAERAFKSNRIOTV--EFLG----SKLKVIGEASFQDNNLRNVMLPDGLKTESEAFT 358
 Db 1236 IQHISSENANDLEGIDSDENLGKNTENMGNTYEEFIKSYNL---ITNYLETVSKEIT 1291
 QY 359 GNPGDEHYNNQVVLRTGPOLATENTYVNPDKSLWRAATPDMDYTKWLEEDFTYQKN 418
 Db 1292 -----YNOI-----ON-KRIPDOKELLKNIENVNKAQSYLDYIK--ENEF--DR 1330
 QY 419 SVTGFPSNKGLQKVRRNKLIEPKOHNGI-----TTEIGD-NAFRNVDFOQSTL-- 466
 Db 1331 IVTHEFKKK-LNTVNDNFKNEYSKVNEGFDNISNSINTVKNSTDENSILNLNQTKEMYAN 1389
 QY 467 -----RKYDLEE----KLPSIRKIGAFAFQSNNLKSFEASBDEEIKEGAFMNN 513
 Db 1390 IVNNTYSYKYEAEINTFNPKLANTLN----IKTNSSGIDLSDKI-KIALSYLD 1442

QY 514 RIGTLIDKDKLIT-----KIGDAAFHINHTIAYAVLP--ESVOETIGRSAFRONGALHLMF- 564
 Db 1443 K-----TEDTLIPFSPQKTTETYKISDYSISILDLIKKSQBLQK---KEQOTIKLIFE 1494
 QY 565 ---IGNKVKTIGEMAFISLNKLESVNLSEQKQKLTIEVQAFSDNALSEVLP-PNLQTRE 620
 Db 1495 NRRILYEVQATNE--LRGTLSDLKYKKEKILSEVKLILHKSNEINKLSCNFONYDTILE 1551
 QY 621 EA-----FKRNLKEVKGSSTSLSQITFNAFDQNDGDKRFGK--KVVRTHNNSHMLADG 672
 Db 1552 SSKYDQVKEKSNNYKQEKBLQD-----FNVTDMEEKFNNDIKVIEELNNYDSSEEN 1605
 QY 673 ERPIIDPDKLSSTM---VDLEKV-LKIIBGLD-----YSTLROTTQ 711
 Db 1606 NNIQSKQKLKELTNKENAELKIDKLTIEKNDLIDKLIETRKNCMLFTHTTLAETKI- 1664
 QY 712 FREMTAGKALLSKSNLROQEOKFLQBAQFFLG-----RVLDKAIAKA 756
 Db 1665 --KLTDSKFLESAT-----KESKEFLKYIGDTSNSLNDIATLQLKYDHLQ-INKY 1713
 QY 757 EKALVTKKATKNGHLLRS-----INKAVLAINNSAIK 789
 Db 1714 VTSKLSDATNDNNLIEKEKEATQAIKNLTKLFTIDSNNDANALHNKIQMVYFENSELH 1773
 QY 790 KA--NVKRLKEKELDLITLVEKGKGLAQATMVQGVYL-----LKTPPLPPEYYIGLNV- 840
 Db 1774 KSIESKOLYKQMHVFKLNLIG-----INGKYFDISKOFDNILQLOQESLTANLND 1825
 QY 841 -----YFDKSGKLTYALDMSTTIGEGOKDAYGNPILNVDEDNEGYHTLA-----V 885
 Db 1826 LKEIGOKISDKKNKFLHALNET-----PIPNFNTLKEIYDIVKKRQIDEI 1872
 QY 886 ATLADYEG---LYTKDILNSSLDKTKAI-----RQIPLAKYHRLGIFQAIRNAAE 933
 Db 1873 ENITSEENENITLYI-DTITKLKEKVQSIINFVTYENDSNIKQH---IQDTNENDVSK 1928
 QY 934 ADRLPKTPKGY---LNEVPNYRKOMEKN-----LKPVDYKTPF 971
 Db 1929 IKESLKTTOQSFOHILINKINGIKAQFYDNNNNNNNNNTISTISQDVNDVKGHISKDLTIE 1988
 QY 972 NKALPNEKVGDRAAKGHTINAETNNSAVTPIRSEQQHLKSOSDVNLQPTSSKNNFIYE 1031
 Db 1989 NELIEIQKSLEDIKNSTYEIRGANNNYNTIRNYVEQQTNKIQNNSNKDEI---DDIIOK 2045
 QY 1032 ILGY 1035
 Db 2046 ILNY 2049

Search completed: April 27, 2004, 11:16:32
 Job time : 34 secs

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 27, 2004, 11:14:07 ; Search time 23 Seconds
(without alignments)

Scoring table: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PC-TUS-COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backflesl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	226.5	4.2	3696	4 US-09-134-001C-5080
2	222	4.1	10182	4 US-09-134-001C-3159
3	201.5	3.8	1786	3 US-08-973-462-8
4	200	3.7	872	3 US-08-851-843A-8
5	200	3.7	872	3 US-08-851-843A-54
6	200	3.7	872	3 US-08-974-549A-221
7	200	3.7	872	3 US-08-854-050-8
8	200	3.7	872	3 US-08-854-050-54
9	200	3.7	872	4 US-09-430-323-8
10	200	3.7	872	4 US-09-430-323-54
11	200	3.7	872	4 US-09-402-181B-221
12	200	3.7	872	4 US-09-721-456-221
13	192.5	3.6	2662	4 US-09-595-684B-31
14	189	3.5	1221	4 US-09-107-532A-3959
15	188	3.5	1312	2 US-08-592-126-148
16	188	3.5	1312	2 US-08-687-080-51
17	188	3.5	1312	4 US-09-168-595-148
18	185	3.5	1972	4 US-08-875-435B-4
19	183.5	3.4	1315	4 US-09-200-650E-5
20	182	3.4	1165	4 US-09-200-650E-7
21	180.5	3.4	840	3 US-08-974-549A-190
22	180.5	3.4	840	4 US-09-402-181B-190
23	180.5	3.4	840	4 US-09-721-456-190
24	179	3.3	3248	1 US-08-353-700-1
25	179	3.3	3248	5 PCT-US9516216-1
26	179	3.3	3418	2 US-08-603-753D-4
27	179	3.3	3418	3 US-09-099-753-4

ALIGNMENTS

RESULT 1

US-09-134-001C-5080

; Sequence 5080, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLO

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134, 001C

; CURRENT FILING DATE: 1998-08-13

; PRIORITY APPLICATION NUMBER: US 60/064, 964

; PRIORITY FILING DATE: 1997-11-08

; PRIORITY APPLICATION NUMBER: US 60/055, 779

; PRIORITY FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5080

; LENGTH: 3696

; TYPE: PRT

; ORGANISM: staphylococcus epidermidis

; US-09-134-001C-5080

Query Match 4.2%; Score 226.5; DB 4; Length 3696;

Best Local Similarity 19.6%; Pred. No. 1.3e-06;

Matches 248; Conservative 186; Mismatches 459; Indels 371; Gaps 54;

Sequence 5080, AP

Sequence 3159, AP

Sequence 8, Appli

Sequence 8, Appli

Sequence 54, Appli

Sequence 221, APP

Sequence 8, Appli

Sequence 54, Appli

Sequence 8, Appli

Sequence 54, Appli

Sequence 221, APP

Sequence 221, APP

Sequence 31, Appli

Sequence 3959, AP

Sequence 148, App

Sequence 51, Appli

Sequence 148, App

Sequence 3959, AP

Sequence 148, App

Sequence 51, Appli

Sequence 190, APP

Sequence 4, Appli

Sequence 5, Appli

Sequence 7, Appli

Sequence 190, APP

Sequence 190, APP

Sequence 190, APP

Sequence 1, Appli

Sequence 1, Appli

Sequence 4, Appli

Sequence 4, Appli

QY 33 SVKQEQTQASASEDDWEE----EDNERKTNVSKE--NSTVDETFVSDLFSDGNSNNSSK 84

Db 1242 NVRKKQEAQTTINDIVQQHQKQSIQNNDATTEEKEVANNLVNASQQNVISKIDNATTNQ 1301

QY 85 TESVVSDFPKQVKAKPEVTOEASNSNDA----SKVEVPKQDTASKKE----TLET 132

Db 1302 IDGIVSDGRQSNIAITPDTSIKRNNAKNDIDIKAADKKKIQRINDATDEEIQEANRKE 1361

QY 133 STWEAKDFVTRGDT--LVGFSK-SGINKLSQ-TSHLVLPSHAADGQ--LTOVASFAFT 185

Db 1362 AKIEAKDNQIQRNSTRDQVNNEAKTNGINKIENITPATVKSARQAVQNKANEQINHQINT 1421

QY 186 PD----KKTAAEYTSRIGE----NGKPSRIDLQDEKELIDE-----EFLNAYVOLTKT 231

Db 1422 PDATNEEKQEAIRVSARVQAQINAETTQGV--KTIKDDAITSLSRINAQVEKES 1479

QY 232 TPNG--YKSTIGQDAFDVN----KNIAEVNLPELETISDYAFAHMSLKVQKL 277

Db 1480 ARNAIEQKATQQTQFINNNNDAEKEVANNLVIATKQKSLDNIN----SLSSNND 1532

QY 278 PDNLKVIGELAFFFQDQIGGKLYPRHLIKLAER----AFKSNRITQFVFLGSKLV 330

Db 1533 VENAKVAG----INETAN--VLPATAVSKAKKDIDQKLAQQINQIQTQHQTATEEK- 1583

QY 331 GEASFQDNLRNVMLPDGLEKISEAFTGNPGDEHYNNQVLRTRTGONPHOL----- 383
 Db 1584 -EAAIQLANQKS-----NEARTAIQ-NEHSNNNGVAQAKSNGIHEIELVMPDAHK 1630
 QY 384 -----ATENTYVNPDKSLWRATPD-----MDYTKWLEEDFTYQKNSVTGFSNKG 428
 Db 1631 KSDAKOSIDNKY-NEQSNTINTTPDATDEEKOKALDKL-K-TAKDAGYNK----- 1678
 QY 429 QKVRRNKOLEIPKOHNGITTEIGDNAFR---NVDFOSKTLRKYDLEEIK-----LPST 479
 Db 1679 DQAQTNQVQSDAKTEAIDTINTIQANVAKKPSARVELDSK-----FEDLKROINATPNA 1732
 QY 480 IRKIGAFAFQSNNLKSFEEAS-----EDLEI-----KEGAFM- 511
 Db 1733 TEEBKQDAIQRNLNGKRDEVKNLQINQDRDNEVEQHKNIGLQELETIHANPTRKSDALQEL 1792
 QY 512 -----NNRIGTLDLKD-----LIKIGDAAFH-----INHI 537
 Db 1793 QTKPFLISQTELINNNKDATNEEKEAKRLLEISKNTITINNQVQDNAKONGMNEI 1852
 QY 538 YAIVLPESVQEGRSFRONGALHLMFI-GNKVKTIGEMA----- 576
 Db 1853 ATIPATIPTIKTDAKTAIDKKAEEQQVTIINGNNDATEEKAEARLVEKAKIEAKSNITNS 1912
 QY 577 -----FLSNKLESV-NLSEQKOLKTIEVQAFSDNALSEWV---LPPNLQTIREAF 623
 Db 1913 DTEREVNGAKTNGLEKINNIQPOSTKTMNAKQEINDKAQEQOLIQINNTPDATEEKEQEA 1972
 QY 624 KR-----NHLKEVKGSSTLSQITNAFFDQNDGDKRFGKKVVVRTHNNSHM 668
 Db 1973 NRVNAGLAQAIQNINAHSTOEVNESKNTSIKSVQPNVKKPTAINSILQEEANNQKT 2032
 QY 669 LADGERFIDPDKLSTMDLEKVLKLTIEGLDYSTLRQTQTOFREMTTAGKALLSKNL 728
 Db 2033 RQ-----LIGNDGNATDEKEAKQQLVQKLNQEQIQUESTQDNQVNDVKAQAITAIKLINANAHK 2092
 QY 729 ROGEKQKFLOEAQFFLGRVLDLKALAKAELVTKKATKNGHLLERSINKAV-----LA 782
 Db 2093 RQ-----DAINTLNLAEKSQSDI--RANQDATEEK--NTAIQSIDDTLA 2134
 QY 783 YNNSAIKKANVKRLEKELDLTDLIVEKGKGLAQATMVGVYLLKTPPLPEYYIGLVNVF 842
 Db 2135 QARNNINGANTNALVDE-----NLEDGKOKLQ----- 2165
 QY 843 DKSGKLIYALDMSDTIGEGOKDAYGNPILINVDEDNE-----GYHTLAVATLADYE 892
 Db 2166 STOKTKOAKADIAQAIQORSTIDONONATTEKEQEAHLRNQETNGVNDRIOAALANON 2225
 QY 893 GLYIKDILNSSLDKTIAKROIPLAKTHRLGIFOAIRNAAEADRLLPKTPK----- 943
 Db 2226 --VTEDEKNNLETIRNVERPIVVKPKANEI--IRKKAEEQOTTLINQNDATLEEKQIA 2279
 QY 944 -GYLINEVPNVRKKOMEKNLKPVDYKTPF-----KALPNEKVGDRAAKG 988
 Db 2280 LGKLEEVKNEALNQVSQAHSNNDVKAENNGIAKISEVHPETIICKRNAQOETEQDAQSQI 2339
 QY 989 HNINA---ETNN-----SVA-----VTPIRSEQQLHKSQSDVN-----LPQTS 1023
 Db 2340 DTINANKSTNEEKAIDRVNVAKIDAINNITNATTQVLVDAKNSGNTSISQILPSTA 2399
 QY 1024 SKNN 1027
 Db 2400 VKTN 2403

RESULT 2
 US-09-134-001C-3159
 ; Sequence 3159, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134 001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIORITY APPLICATION NUMBER: US 60/064, 964
 ; PRIORITY FILING DATE: 1997-11-08
 ; PRIORITY APPLICATION NUMBER: US 60/055, 779
 ; PRIORITY FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3159
 ; LENGTH: 10182
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-3159

Query Match 4.1%; Score 222; DB 4; Length 10182;
 Best Local Similarity 19.7%; Pred. No. 1.2e-05; Mismatches 435; Indels 324; Gaps 58;
 Matches 236; Conservative 200; Mismatches 435; Indels 324; Gaps 58;
 Db 8438 QREVMINTNTNATTREKVKVNLQQA-LDKAMETLQQVVAHKNNITLNDSKYINEDSKY 8495
 QY 29 EREESVKQEQTQSASEDDWFEEDNERKTKNSKENSTVDETV---SDIFSDGNSNNSSSKT 85
 Db 8496 QOYQDRVIADEQ----LINCTNPTLEPKVVDIVKDNVLANEKILFGAELSYDKS 8548
 QY 136 EAKDFVTRGDTIVGFSKSGINKLQSTSHLVLBESHAADGTQLTQVASFAFTPDKKTAIEY 195
 Db 8549 NANDEIKHMYLNNNAQKSQIKDM-----SHAALRTEVKQL----- 8584
 QY 196 TSRIGENGKPSRLDIDQKETIDEG---EIFNAYQLTKLIPNGYKSTQD----- 242
 Db 8585 -----LQQAKILDEAMKSLEDKIQVWITDTLPN-YTEASEDKKEKVQDQTV 8629
 QY 243 ---AFVDNKNIAEVNL-----ESLETISYAFAHMSLKQVQLPDNLKVIGELAFFDN 292
 Db 8630 SHAQAIIDKINGSNVSLDQVRQALEQLTOQASHENLDGQDRVEEAKVHAN-QTIDQIHLNS 8688
 QY 293 --QIGGKLYLPRHLIKIARAFKNSRIQTVFELGSKLKVIGEASFQDNRLRNVMLPDGLE 350
 Db 8689 LQQQTAKESV-KNATKLEELATVSNNAQALN-----KVMGKLEQFINHADSVENSIDNYR 8741
 QY 351 KIESEAFTG-NPGDEHYNNQVLRTRGONP----HOLATENTVNPDKSLWRATP-DM 403
 Db 8742 QADDKLTIAVEALEH--GQDIQKTNTQNETKQALQQLIYAETSLNGFERINHARPRAL 8799
 QY 404 DYTWKLEEDFTYQKNSVTGFSNKGQLQVRRNKNLEIPKQH--NGITITEI-GDNAFRNV 459
 Db 8800 EYIKSLEKINNAQKSALE-----DKVTOSHDL-LELEHTVNEGTLNLDIMGELANAV 8851
 QY 460 DFFQSKTLRKYDLEEIKLPSKSTIRKIGAFAFQSNNL-KSPEASEDLEEIKEGAFMN-NRIGT 517
 Db 8852 NYTAPT-----KASINYINADNLRKNFQAINNARDALNKTQGQNLDFNAIDT 8900
 QY 518 LDLDKDKLLIKIGDAAPHINHYAI----VLPESVQEGRSF--RONGALHLMFIGNKV 570
 Db 8901 --FKDIDFKTDALNGIERLTAKSKAELTDSLKFINKAQFTHANDEIMNTNSTAQLSR 8958
 QY 571 TIGEMAFLSNKLESVN-SEQKOLKTIEVQAFS-----DNLASEVVLPPNQLT 617
 Db 8959 IVNQAFDINDAMKS-LRDELNNQARPVQASSNYINSDEBDIKQOFDHALSNA---RKV 9011
 QY 618 IREEAFKRNLKEVKGSSTLSQITNAFFDQNDGDKRFGK--KVVVRTHNNSHMLADGER 674
 Db 9012 LAKENGKNLDEKQIQC---LKQVIEDTKDALNGIQRISKAKAKAIQYVQSLY-INDAQ 9067
 QY 675 FTID----PDKLSS-----TMVDLEKVLK----IIEGLDYSTLRQTTQ 711
 Db 9068 HIAENNIHNSDDSSLANTLSKASDLDNAMKDLRDTIESNSTSVPSVNTINADKNLQIE 9127
 QY 712 FREMTTAGKALLSKS-----NLROGEKQKFLQEAQFFLGRVLD 750
 Db 9128 FDEALQQASATSSKTSIENPATIEEVGLGLSQAIYDTKNAALNGE-QRLATEKSQDKLIKGL 9186

QY 751 KAIKAKAERKALVTKKATKNGHLLERS-----IN--KAVLAYNN 785
 Db 9187 KDLNKAQLEDVTNKVNSANTLTTELQSTLNDKMKLURDKJKTLLVNPVKASLNTVN 9246
 QY 786 S-----AIKKA-----NVKLEKELDLTLVEGKGPLAQATMVQ 820
 Db 9247 ADYNLKRQFNKALKEAKGVLNKNNSGTNVNINDIQHLLTQIDNAQDOLNGERRIKEHQKS 9306
 QY 821 GYVLLKTPPLPEYYIGLN-----VYFKSGKL-----YALDMSDTIGEQOKDAY 866
 Db 867 GN-----PILNVEDDENEGHTLAVATLADYEGLYIKDILNSLDKIKAIRO-----IP 914
 QY 9307 EVFIK-----ELDILNNAQKAATINQIRASKDIKINQIVDNAELNDAM-OGLKEHV 9359
 Db 9360 AOLTATPKONIEYLNADEDHKLOYDYAI-----NLANNVLDKENGTNKDANI 9407
 QY 915 LAKYHRIGIFQAIRNAAEADRLLPKTPKGYLNEVNPYRKOMEKNLKPVDYKTPFNA 974
 Db 9408 IGMIONMDARALLING---TERLKDAQTKAH-NDIKDTLKRODE---IEHANATNSK 9459
 QY 975 LPNEKUDGSDRAAKG-HNINAETTNSV-----ANTPIRSEQQLHKSQSDVN 1018
 Db 9460 AQAKQMVNEARKALSNINDATSNLNUQAKDEGOSAIEHAD-ELPKAKLDAN 9513
 RESULT 3
 US-08-973-462-8
 ; Sequence 8, Application US/08973462B
 ; Patent No. 6191270
 ; GENERAL INFORMATION:
 ; APPLICANT: DAUBERSIES, PIERRE
 ; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
 ; FILE REFERENCE: 0660-0125-0 PCT
 ; CURRENT APPLICATION NUMBER: US/08/973,462B
 ; EARLIER APPLICATION NUMBER: PCT/FR96/00894
 ; EARLIER FILING DATE: 1996-06-12
 ; EARLIER APPLICATION NUMBER: FR 95/07007
 ; EARLIER FILING DATE: 1995-06-13
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 8
 ; LENGTH: 1786
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
 ; US-08-973-462-8

Query Match 3.8%; Score 201.5; DB 3; Length 1786;
 Best Local Similarity 20.0%; Pred. No. 2.7e-05;
 Matches 228; Conservative 176; Mismatches 433; Indels 305; Gaps 48;

QY 31 EESVKQKOTOSASEDDWFEEDNERKTNVSK---ENSTVDETVSDLJFSDGNSNNSSKTES 87
 Db 756 EESVEENTVEESVAE-----NVEESVAEN--VEESVAE----NVEESVAPTE 796

QY 88 VVSDPKQVKPKPEVTOE----ASNSSNDASKVKEVPKQD TASKETLETSTWEAKDFV- 141
 Db 797 EIVAPSVEESVAPSVSEESVAENVATNLSDNLLSLLGIGETEEIKDSILNEEEVKENV 856

QY 142 -----TRGDTLVGFSKGINKLSQTSHLVLPSHAADGTLQVASFAFPDPDKTAI 192
 Db 857 TTILENVEETTAESTVTTFSNI-----LEEIQENTITND---TI 891
 QY 193 AEYTSRLGENGKPSRIDIQ---KEIIDEGE-----IFNAYQLTKLTIPI 234
 Db 892 EEKLEELHENVLSAALENTOSEEKEKEVIDVIEVKEEVATFLEETVEQEEKSANTITE 951

QY 235 GYKSIGODAFVDNKNIAEVNLPELESLETISDYFAHMSLKQVKLPDNLKVG----- 286

Db 952 IFENLEENAVESNENVAE-NLEKLNFTVNTVLD-----KVEETVETSGESLENNEMD 1003
 QY 287 LAFFDNOIGGKLYLPRHLIKLAAERAFKSN-RIQTVEFLGSKLKVIGEASFQDN--NLRNV 343
 Db 1004 KAFFSEIFDNVKGQIENNLITGMFRSIEITSIVQSEEKVDLNENV--SSILDNIENMKEG 1061
 QY 344 MLPDGLEKIESEAFTGNPGDEHYNNQVVLRTTGONPHQLATENTYVNPDKSLWRAIPDM 403
 Db 1062 LL-NKLENISISTEGVQETVTEH-----VEQNVTVDVD----VPAM 1096
 QY 404 DYTWKLEEDETYOKNSVTGFSNK--GLOKVRNRNKNLEIPKOHNGITITEIGDNAFRNDF 461
 Db 1097 -----KDQFLGILNEAGGLKEMFFNLEDVFKSESDVITTEEIKDEPVQK-EV 1142
 QY 462 QSKTLRKYD-----LEETIKLPSITRKIGA---FAFOSNNLKSFEASED----- 501
 Db 1143 EKETVSIIIBMEENIVDVLBEEKEDLTKMIDAVEESIEISSDSKETESTIKDEKDVS 1202
 QY 502 -LEIKEGAEMNMRIGTLKDKLJIKIGDAAFFHINHTIATLVPESVQETGRSAFRQNGAL 560
 Db 561 HLMFIGNKVKTIGEMAFLSNKLESVNLSEQKQKLTIEVQAFSDNALESEVVLPPNQTI 620
 QY 1256 DLIKDMEKREL-EKALSEDSEKELTDAKDTLEKVIEEHDITTTDEVV---ELKOVEE 1311
 Db 1203 VVEEVQDNDMDESVEKVELKNNMEELMKDAWEINDITSKL-EBTQELNEV----EA 1255
 QY 621 EAFKR-----NHLKEVKGSSTL-SQI-TFNAFDQDGDKRFGKVVVRH----- 663
 Db 1312 DKIEKVSSDLRQLEEDILKEVKEIKELESETILEDYKELKTIETDILEEKKEIEKDHFPEKFE 1371
 QY 64 NNSHMLADGERFLIIDPDKLSSTMVDEKVLK-----IIRG----- 698
 Db 1372 EEAEEIKDLEADIL--KEVSSLVEEKKLEEVEHLEKEVHEHISGDAHKGLEDDLEB 1429
 QY 699 -----LDYSTLROTTQFREMTTAGKALLSKSNLROGEKQKFLQ 738
 Db 1430 VDDIKGSILDMKLKGDMELGDMDKESLEDVT-TKLGERVESLKDVLSSALGMDEEQMKTRK 1488
 QY 739 EAQFLGLRVDLKDIAKAE-----KALVTKKATK-----NGHLLERSINKAV 780
 Db 1489 KAO---RPKLEEVLLKEEVKEEPKKITKKVRFEDIKDEPKDEIVEVEMKDEDIEBV 1544
 QY 781 LAYNNSAIKKANVKRLEKEBLDILTLVEGKGPLAQATM----VOGYVLLKTPL--PLPE 833
 Db 1545 EEDIEDEBKVDEDIDEDID--EDIGEDKDEVIDLIVQKEKRIEKVAKKKLEKKVE 1602
 QY 834 YYIGIINYFDSKGKLIVYALDMSDTIGEQOKDAYGNPILNVDENEGYHT----- 882
 Db 1603 GVGGLKKHVDEVMKVYQKIDKEVDEKVSKEKNDVTNLQKQNDFFSKVKNFVKKV 1662
 QY 883 -----LAVTLADY-EGLYIKDI-----LNSSLDK-IKAIRQIPLAKYHRIG 922
 Db 1663 FAAPFISAVIAFASYVVGFFPSLFLSSCVTIASTYLLSKVDKTINQNKERPFYF----- 1718
 QY 923 IFQAIRNAAEADRLLPKTPKGYLNEVNPYRKOMEKNLKPVDYKTPFENKALPNEKVGD 982
 Db 1719 VFDFIKNLKYLQQMKEKFSKEKNINNVIEWTNAKAEGKGNVQVTNKTEKTTKVDKNNKVK 1778
 QY 983 DR 984
 Db 1779 KR 1780

RESULT 4
 US-08-851-843A-8
 ; Sequence 8, Application US/08851843A
 ; Patent No. 6093809
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru
 ; APPLICANT: Chapman, Karen B.
 ; APPLICANT: Morin, Gregg B.

APPLICANT: Harley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6093809el Telomerase
 NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDBEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-851-843A-B

Query Match 3.7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TTVSVVTVYQEVVYGLERESVKEQOTQASADEDDWFEEDNERK- 55
 Db 17 TNLDFVLQNLLEVYKQSQIEHYKTTQQQIKEEDIKLKFKNQDQDGNSGNPDDDEENNSNQ 76
 QY 56 TNVSKENSTVDETDVSDLFDSDGNNSNSSSKTESVSDPKQVKAKPEVTQEASN 108
 Db 77 QELLRRVNOIKQQVOLIKKVGSKVEKDLNLN-EDEN 111
 QY 109 SSNDASKVEVKQDKTAKKETIETSTWEA--KDFVTRGDTIVGFSKSG-- 154
 Db 112 KKNGLSEQQV-----KEBQIERTTIEQVKYQNLVNMDFQLDLINESGGRRHRETDY 164
 QY 155 -INKLSQTSHVNLPSHAADGTTQVASEAFTPDKKTAIAEYTSRIGENGKPSRIDQK 213
 Db 165 DTEKWFEFISH-----DQKNYVSIYA--NOKTSYCWHLKDFNKNNYDHLNSTN 211
 QY 214 EIDEGEIEF--NAVQLTKLITIPNGYKSIQGDAFVDNK-----NIAEV- 253
 Db 212 RLETEAEEFYAFDDFSQTIKLT-NNSYOTVNIDVNFDNNLCILALLRFLLSLERFNLNIR 270
 QY 254 -----NLPESELTISDYFAHMSLIKQVKLPDNLKVGELAFFDNQIGGK---- 297

RESULT 5
 US-08-851-843A-54
 Sequence 54, Application US/08851843A
 Patent No. 6093809

GENERAL INFORMATION:
 APPLICANT: Cech, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Harley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6093809el Telomerase
 NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

Db 271 SSYTRNQYNFPEKIGELLETIFAVVFSRHRQGITHQVPCFAQYLVNNSQISVQKDSQLO 330
 QY 298 LY-----LPRHLIKAERAFKSNSRIQTVFELGSKLKVGAEAS 334
 Db 331 VYSFSTDILKLVDTNKVQDYFKEFPR-LTHVSQQAIPVSATNAVENINVLKVKHA- 388
 QY 335 FQDNNLRNVMLPDGLEKIESEAFTGNGPDEHYNNQVVLRTGQNPHQLATENTYVNPDK 394
 Db 389 ---NLNTIVSIP-----TQFNEDFYFVNQHILKLEFGLEPNILTKQKL----E 428
 QY 395 SLWRATPDMYTKWLEDF-TY--QKNSVTGFSNKG--LKQVRNMKOLEIPKQHNGIT 449
 Db 429 NLLSIKQSKNLUKFLRLNFYTYVAQETSRKQILQATTIKNLKNNKQDEETPETKDETPS 488
 QY 450 EIGDNAFRNVDFQSKTRKYLDEEIKLPST--IRKIGAFAFQSNMKSFEAS-EDLEE 504
 Db 489 E-STSGMKEFDHLSELTELEDF-SVNLQATOEYDSLHKLIRSTNUKKFLSYKEM 546
 QY 505 IKEGAF--MNRRIGTL-DLKDQKLIKIGDAFAHINHNIYAIYLPESV-----QBT--G 550
 Db 547 SKMDTFLDILKNIYETLNLNKRCVNSP--HGNTISYELTINKDSTFYKFLTLNQELQHA 604
 QY 551 RSAFRONGAIJHLMFIGNKVKTIGEMAFLSNKSLSV-----NLSEOKOLKTIEVQAFS 602
 Db 605 KYTFKQN----EFQFNVKS--AKIESSLESLEDIDSLCKSTIASCKNLQNV-----650
 QY 603 DNALSEVVLPPNLQTIETEEAFKRNHL--KEVKGSTSLSQTFNA-FDQNDGDKRKGK 658
 Db 651 -NITASLLYPPNNIQ--KNPFNKPNNLFFKQFPEQLKNLNEVSINCIDQ-----695
 QY 659 VVRTHNNSHMLADGERFTIDPDKLSSSTMVDLKVLKTIIEGLDYSTLRQTTQ-----709
 Db 696 -----HILNSISEFLEKNKKIKAFILKRYVILQY--LDYTKFLKFTLQLQBLNQV 745
 QY 710 --TQFREMTTAGKALLSKSNLROGEKOK-----FLOEAQFFLGRVLDKAIA---KA 756
 Db 746 INQQLEELTVSE---VHKQVWENHKQAFYBPLCEFIKEKQSSQTLQLIDFDQNTVSDDSI 801
 QY 757 EKALVTKKATKNGHLLERSINKAVLAVNNSAIKKANVKRLEKEILDLLTDLVEGKGPLAQ 816
 Db 802 KKILESESKYHHLRINPSQ----SSSLIKSEN----EIQEELLKACDEKGVLVKA 851
 QY 817 TMVQGVYLLKTPPLPPEYYIGLNVYFD 843
 Db 852 -----YKKFPLCLP----TGTYYD 866

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 ; ; ;
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 ; ; ;
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 ; ; ;
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 TELECOMMUNICATION INFORMATION:
 REGISTRATION NUMBER: 36,429
 INFORMATION FOR SEQ ID NO: 54:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: peptide
 ; ;
 US-08-851-843A-54

Query Match 3 7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05; Mismatches 350; Indels 294; Gaps 48;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

Qy 14 TRVSVWVTVSQEVYKLE-----REESVK-----QEQTOSASEDDWFEEEDNERK- 55
 Db 17 TNLDFVQLQNLLEVYKQSIEHYKTTQQQIKEEDKLKILKFKNQDDQDGNSGNDDEENNSNQ 76
 Qy 56 -----TNVSKENSTVDETVSDLFSDGNSNNSSSKTESVVSDPKQVPKAKPEVTOEASN 108
 Db 77 QELLRRVNOIKOQVQLIKKVGSKVEKDNLNN-----EDEN 111
 Qy 109 SSNDASKVVEPKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKSG--- 154
 Db 112 KRNGLSEQQV-----KEEQLRTITEEOKVQYQNLVFNMDYQDLDLNEGGHRRHRRRTDY 164
 Qy 155 -INKLSQTSHLVLVLP SHAADGTOQI TQVASFAFTPDKKTAI AETSRSRIGENGKPSRSLIDQK 213
 Db 165 DTEKWFEEISH-----DQKNNYVSIYA---NQKTSYCWWLKDYNKNNDHNVSN 211
 Qy 214 EIIDEGEIFF---NAYQLTKLIPNGYKSIGDAFVDNK-----NIAEV- 253
 Db 212 RLETEAEFYAFDDFSTQTIKLT-NNSYQTVNIDVNFDNMCILALLRFLJSLERFNILNIR 270
 Qy 254 -----NLPESELETISDYAFAHMSLKVQKLPDNLKVIGELAFFDNOQIGK----- 297
 Db 271 SSYTRNQYNFKEKIGELLETIFAVVFSHRHLQHQVPCAFQYLVNNSSQISVKSQDQLQ 330
 Qy 298 LY-----LPRHLIKLAERAFKSNRIQTEFLGSKLKVIGEAS 334
 Db 331 VYSFSTDLLKLVDTNKQDYFKEFPR-LTHVSSQQAIPVSATNAVENLNVLKKVKA- 388
 Qy 335 FQDNLLRNMLPDLGKIESEAFTGNGPQDHEHYNQVVLRTRTGONPHOLATENTYVNPDX 394
 Db 389 ---NLNLVSI-----TOFNFDYFVNLQHLKLEFGLPMLTKQKL-----E 428
 Qy 395 SLWRATPMDYTKWLEEDF-TY--OKNSVTGFSNKG--LQKVRNKNLETPKQHNGITTT 449
 Db 429 NLLSISIKQSKNLKFLRLNFYTYVAQETSRKQILKQATTIKLNKNKNQEEETPETKDETPS 488

RESULT 6
 US-08-974-549A-221
 ; Sequence 221, Application US/08974549A
 ; Patent No. 6116178
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru
 ; APPLICANT: Chapman, Karen B.
 ; APPLICANT: Morin, Gregg B.
 ; APPLICANT: Harley, Calvin B.
 ; APPLICANT: Andrews, William H.
 ; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
 ; NUMBER OF SEQUENCES: 727
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/974,549A
 ; FILING DATE: 01-OCT-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 18-APR-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/846,017
 ; FILING DATE: 25-APR-1997
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-00261
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-549A-221

REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 54:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: peptide
 /US-08-854-050-54

Query Match 3.7%; Score 200; DB 3; Length 872;

Best Local Similarity 20.3%; Pred. No. 1.2e-05; Mismatches 350; Indels 294; Gaps 48; Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TTVSVVTVSQEVYGLR-----RESVK-----QEQTQSASEDDWEEEDNERK- 55
 Db 17 TNLDLFLVQNLLEVYKSQLIEHYKTOQQQIKEEDLKLKEFKNQDQDGNSGNDDEENNSNQ 76
 QY 56 -----TIVSKENSTVDETVSDLFLSDGNSNNSSSKTESVSDPKQVPAKEBTQEASN 108
 Db 77 QELLRRVNVQIKQQVQLIKVGSKVEKDLNN-----EDEN 111
 QY 109 SSNDASKVVEVPKQDTASKKETLETSTWEA--KDFVTRGDTLVGFSKSG----- 154
 Db 112 KKNGLSEQQV-----KEEQLRTITEEQVKYQNLVFNMDYQOLDLNEBGGHRRRETDY 164
 QY 155 -INKLSQTSHTSHLVLPLSSHAADGTOITQVASFAFTPDKKTATAEYTSRLGENGKPSRLDIDQK 213
 Db 165 DTEKWFETISH-----DQKNTSYVSIY-----NQKTSYCWNLKDYNFKNNYDHLNVSIN 211
 QY 214 ELLDEGEIF---NAVQLTKLITPINGYKSTGQDAFVDNK-----NTIAEV- 253
 Db 212 RLETEAEFYAPDDFSQTIKLT-NNSYQTVNIDVNFDNNLCLALLRFLLSLERFNILNIR 270
 QY 254 -----NLPESELETISDYFAHMSLISKQVKLPDNLKVGELAFFDNOIGK----- 297
 Db 271 SSYTRNQYNEFKIGELLETIFAVVWFSHRHLQGIHLQVPCFAQYLVNNSSQISVKDSQLQ 330
 QY 298 LY-----LPRHLIKAERAFFKSNRIOQTVEFLGSKLKVIGEAS 334
 Db 331 VYFSTDLKLVDTNKVQDYFKFLQEFPR-LTHVQSQQAIPVSAATNAVENLNVLKKVKHA- 388
 QY 335 FQDNNLNRNVMLPDGIEKIESEAFTGNPGDEHYNNQVVLRTRTGQNPHQLATENTYVNPDK 394
 Db 389 ---NLNLVSI-----TOFNFDYFVNLOHLKLEFGLEPNLITKOKL-----E 428
 QY 395 SLWRATPDMYTRKLEEDF-TY--OKNSVTGFSNKG--LQVRRNKNLIPKQHNGITIT 449
 Db 429 NLLSIQSKNLUKFLRLNFYTYVAQETSRKQILQATTIKLNKNKQNEETPETKDETSP 488
 QY 450 EIGDNAPRNVDFOSKTLRKYDLEEIKLPST--IRKIGAFQSNNLKSPEAS--EDLEE 504
 Db 489 E-STSGMKFFDHLSLETELEDF-SVNLOQATOELYDLSLHKLIRSTNLKKEKSYKEMEK 546
 QY 505 IKGEGAF--MNRIGTL-DLKDKLJIKIGDAAPHINHYATIVLPESV-----QEI--G 550
 Db 547 SKMDTFLKNTIYETTNNLKRCSVNISNP--HGNISYELTINKDSTFYKEKLTNLQELQHA 604
 QY 551 RSAFRQNGALHIMTIGNKVKTIGEMAFLSNKLESV-----NLSEQOKOLKTIEVQAFS 602
 Db 605 KVTFKQN-----EFQFNNVKS--AKIESSSLSEDIDSLCKSISACKNLQNV----- 650
 QY 603 DNALSEVVLPPNLQTIREEAFKRNL---KEVKGSSTLSQITFNA-FDQNDGDKRFGKKV 658
 Db 651 -NIASLILYPNNIQ--KNPFPNKPNLFFKQFEQLKLNENVSINCILDQ----- 695
 QY 659 VVRTHNNSHMLADGERFIIDPDKLSTMVDLEKVLKLTIEGLDYSTLROTQ----- 709
 Db 696 -----HILNSISEFLEKNKKIKAFILKRYVLLQY--LDYTKLFTLQLPELNQWY 745

RESULT 9
 US-09-430-323-8
 Sequence 8, Application US/09430323
 Patent No. 6309867
 GENERAL INFORMATION:
 APPLICANT: Cech, Thomas R.
 Lingner, Joachim
 Nakamura, Toru
 Chapman, Karen B.
 Morin, Gregg B.
 Harley, Calvin
 Andrews, William H.
 TITLE OF INVENTION: No. 6309867el Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/430,323
 FILING DATE: 29-Oct-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/854,050
 FILING DATE: 09-MAY-1997
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-09-430-323-8

Query Match 3.7%; Score 200; DB 4; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;

Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48; QY 14 TVSVWVTVSQEVYGLR-----REESVK-----OEOTQSASEDDWFEEDNERK- 55
 Db 17 TNLDLDFVLQNLLEVYKSQIEHYKTQQQIKEEDEKLKLFKNQDQDGNSGNDDEENNSNQ 76
 QY 56 -----TIVSKENSTVDETDVSLFSDGNSNNSSKTESVVSDFPKQVKPKAKPEVTOEASN 108
 Db 77 QBLRRVNQIKQVQLIKKVGSKVEKDNLN-----EDEN 111
 QY 109 SSNDASKVEVPKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKG----- 154
 Db 112 KKNGLSEQQV-----KEEQLRTITEBQVKQNLVFNMDYQLDLINESGGHRRRRETDY 164
 QY 155 -INKLSQTSHLVLPLSHAADGTLTQVASFAFTPDKKTAAEYTSRLGENGKPSRSLIDQK 213
 Db 165 DTEKWFEBISH-----DQKNYVSIYA---NQKTSYCWMLKQYFNKNNYDHINVSIN 211
 QY 214 EIDEGEIF---NAYQLTKLTIPNGYKSIGQDAFVDNK-----NIAEV- 253
 Db 212 RLETEAEIFYAFDDFSOTIKL-NNSYQTVNIDVNFDNNTCLIAIRFLLSLERFNILNIR 270
 QY 254 -----NLPELETISDYAFAHMSLKQVKLPDNLKVIGELAFFFNDQIGK----- 297
 Db 271 SSYTRNQXNFKEKIGEELLETIFAVVFSHRHIGQIHLQVPCAFQYI-VNSSSQISVTKDSQ 330
 QY 298 LY-----LPRHLIKAERAFKSNRRIQTVEFLGSKLKVIGEAS 334
 Db 331 VYSFSTDILKLVDTNKVQDYFKELOFPR-LTRVSQAAIPVSATNAVENINVLKKVKA- 388
 QY 335 FQDNMLRNVMLPDGLEKIESEAFTGNPGDHEHNNQVVLRTGONPHOLATENTYVNPDK 394
 Db 389 ---NLNLVSI-----TQFNFDFFYFVNQNLQHLKLEFGLEPNLTKQKL-----E 428
 QY 395 SLWRATPDMYTKWLEEDF-TY---OKNSVTGFSNKG---LQKVRMRNKLIEIPKQHNGITR 449
 Db 429 NLLSIKOSKNLKFRLNFTYVAQETSRKQILKOATTIKNLKQKQEEETPETKDETPE 488
 QY 450 EIGDNAFRNVDQSKTLRKYDLEEIKLPST---IRKIGAFAFQSNNLKSFEAS---EDLE 504
 Db 489 E-STSGMKFFDHLSELTBELEDF-SVNLOQATQEYIDSLHLKLLIRSTNLKKFKLISKYEMEK 546
 QY 505 IKGAF--MNRIGTL-DLKDKLIKIGDAFAFHINHYAIVLPESY-----QET--G 550
 Db 547 SKMDTFIDLKNIYETLNLKRCVSNISNP--HGNISYELTNKOSTFYKFKLTLNQELQHA 604
 QY 551 RSAFRONGALHLMFIGNKVKTIGEMAFLSNKLESV-----NISEQKOLKITEVQAFS 602
 Db 605 KYTFKQN-----EFOFNWKS---AKIESSLESLEDIDSICKSIAASKNLQNV----- 650
 QY 603 DNALSEVVLPPNLTQTIREEAFKRNL---KEVKGSSTLSQITFNA-FDQNDGDKRFGKKV 658
 Db 651 -NIIASLJLXPNNIO---KNPENKPNLLPFPROFEQLKNTLEVSINCILDQ----- 695
 QY 659 VVRTHNNSHMLADGERFIIDPKLISSTMVDEKVKIIEGLDYSTLROTT----- 709
 Db 696 -----HLNSISEFLEKNKIKAKFTLKRYYLQY---LDTXKLFKTLQOLPELNLQY 745
 QY 710 --TQFREMTTAGKALLSKSNLROGEKQK-----FLOEAQFFLGRVLDKIA---KA 756
 Db 746 INQQLLELTIVSE---VHKQVWENHKQKAFYEPLECEFLKESOTLQLIDFDONTVSDDSI 801
 QY 757 EKALVTKKATKNGHLERSINKAVLAVNNSAIKKANVKRLEKEDLTLDVEGKGLAQA 816
 Db 802 KKILESISESKYHHLRINPSO---SSSLIKSEN---EETOELLKACDEKGVLVKA 851
 QY 817 TMWQGVTLKTPPLPBEYIQLNVYFD 843
 Db 852 -----YYKFPCLLP-----TGTYD 866
 RESULT 10
 US-09-430-323-54

Sequence 54, Application US/09430323
 ; Patent No. 6309867
 ; GENERAL INFORMATION:
 ; APPLICANT: Czech, Thomas R.
 ; Lingner, Joachim
 ; Nakamura, Toru
 ; Chapman, Karen B.
 ; Morin, Gregg B.
 ; Harley, Calvin
 ; Andrews, William H.
 ; TITLE OF INVENTION: No. 6309867 Telomerase
 ; NUMBER OF SEQUENCES: 225
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/430,323
 ; FILING DATE: 29-Oct-1999
 ; CLASSIFICATION: <Unknown>
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/854,050
 ; FILING DATE: 09-MAY-1997
 ; APPLICATION NUMBER: US 08/844,419
 ; FILING DATE: 18-APR-1997
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 06-MAY-1997
 ; APPLICATION NUMBER: US 08/846,017
 ; FILING DATE: 25-APR-1997
 ; APPLICATION NUMBER: US 08/844,419
 ; FILING DATE: 18-APR-1997
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 01-OCT-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Apple, Randolph T.
 ; REGISTRATION NUMBER: 36,429
 ; REFERENCE/DOCKET NUMBER: 015389-002930US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 54:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 872 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPLOGY: not relevant
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 54:
 ; US-09-430-323-54
 ;
 ; Query Match 3.7%; Score 200; DB 4; Length 872;
 ; Best Local Similarity 20.3%; Pred. No. 1.2e-05;
 ; Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;
 QY 14 TVSVWVTVSQEVYGLR-----REESVK-----OEOTQSASEDDWFEEDNERK- 55
 Db 17 TNLDLDFVLQNLLEVYKSQIEHYKTQQQIKEEDEKLKLFKNQDQDGNSGNDDEENNSNQ 76
 QY 56 -----TIVSKENSTVDETDVSLFSDGNSNNSSKTESVVSDFPKQVKPKAKPEVTOEASN 108
 Db 57 QELLRRVNQIKQVQLIKKVGSKVEKDNLN-----EDEN 111
 QY 77 QELLRRVNQIKQVQLIKKVGSKVEKDNLN-----EDEN 111
 Db 109 SSNDASKVEVPKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKG----- 154
 QY 112 KKNGLSEQQV-----KEEQLRTITEEQQVKQNLVFNMDYQLDLINESGGHRRRRETDY 164
 Db 155 -INKLSQTSHLVLPLSHAADGTLTQVASFAFTPDKKTAAEYTSRLGENGKPSRSLIDQK 213

Db 271 SSYTRNQYNEFEKIGELLETIFAVVVFSHRHLOGIHLQVPCAEFOYLNSSSQISVKSQDQLQ 330

QY LY-----LPRHLIKLAERAFKSNRQIQTVEFLGSKLKVIGEAS 334

Db 331 VYSFSTDKLUVDNKVQDYFKEFQFPR-LTHVSQLQAPVSATNAVENLNVLKKVKA- 388

QY 335 FQDNNLRLNMLPDGLEKIESEAFTGNPGDEHYNQVVLRTTGQNPOLATENTYVNPDK 394

Db 389 ----NLNLTVSIP-----TQFNFDFFYFVNQHLKLEFGLFELNLTQKL----E 428

QY 395 SLWRATPDMYDFTKMLEEDF-TY--QKNSVTGFSNKG--LQKVRNRKNLIEPKOHNGITIT 449

Db 429 NLLSIKQSKNLUFLRNFYTTVAQETSRSKQILKQATTIKNLKNNKQEEETPDKDTPS 488

QY 450 EIGDNAFRNVDQSKTLRKYDLEEEIKLPST--IRKIGAFAFQSNNLKSFEAS--EDLEE 504

Db 489 E-STSGMKFFDHLSELTELEDF-SVNLIQATOBIYDLSHLKLLIRSTNLKKFKLISKYMEMX 546

QY 505 IKGAF--MNRIGTL-DLKDKLKLIGDAFAFHINHITYAVLPESV-----QEI--G 550

Db 547 SKMDTFIDLKNIYETNLNLRCSVNTISNP--HGNISYELTNKDSTFYKFKLTNLQELQHA 604

QY 551 RSAFRONGALHLMFIGNKVKTIGEMAFLSNKLESV-----NLSSEQKQKLTIEVQAFS 602

Db 605 KYTFKQN----EFQFNNVKS---AKIESSLESLEDIDSLSLCKSIAASCCKNLQNV---- 650

QY 603 DNALSEVVLPPNLIOTIREEAFKRNHL--KEVKGSSTLSQITENA-FDQNDGDKRFGKKY 658

Db 651 -NIIASLLYIPNNIQ--KNPFNPKNPLFFKQFQLKNLLENVSINCILDQ----- 695

QY 659 VVRTHNNSHMLADGERFILDPDKLSSTMVDLEKVKLIEGLDYSLTRQTO----- 709

Db 696 -----HILNSISEFLEKNNKIKAFILKRYLLQY--LDYTRKFKTLQQLPELNOVY 745

QY 710 --TQFREMTTAGKALLSKSNLROGEKQK--FLQEAQFFLGRVLDKAIA--KA 756

Db 746 INQQLEELITVSE---VHKQWBNHKOKAFYPLCEEFFIKESSOTLQLIDEPONTVSDDSI 801

QY 757 EKALVTKKATKNGHLERSINKAVLAVNNSAIKKANVKRLEKELDLTLVEGKGPLAQ 816

Db 802 KKILESISESKVHHLRLNPSQ----SSSLIKSEN----EEIQLKACDEKGVLVKA 851

QY 817 TMWQGVVLLKTPPLPIPEYYIGINVFD 843

Db 852 -----YKFLCLP----TCTYYD 866

RESULT 12

US-09-721-456-221

Sequence 221, Application US/09721456

Patent No. 6617110

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

Lingner, Joachim

Nakamura, Toru

Chapman, Karen B.

Morin, Gregg B.

Harley, Calvin B.

Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.3.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/721,456

FILING DATE: 22-No. 6617110-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US 08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US 08/911,312

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

APPLICATION NUMBER: WO PCT/US97/17885

Page 11

QY 298 LY-----LPRHLIKLAERAFKSNRQIQTVEFLGSKLKVIGEAS 330

Db 165 DTEKWFISH-----DQKNYVSIY--NQKTSYCWNLKDFFNKNYDHLNVSIN 211

Qy 214 EIDEGEIF--NAVOLTKTIPNGYKSIGQDAFDNK----NIAEV- 253

Db 212 RLTEAEFYAFDDFSQTIKLT-NNSYQTVNIDVNFDNLICLALLRFLLISLERFLNIR 270

Qy 254 -----NLPRESLETISDYAFAHMSLKVQLPDLKVLVIGELAFFDNQIGGX---- 297

Db 271 SSYTRNQYNEFEKIGELLETIFAVVVFSHRHLOGIHLQVPCAEFOYLNSSSQISVKSQDQLQ 330

Qy 298 LY-----LPRHLIKLAERAFKSNRQIQTVEFLGSKLKVIGEAS 334

```

Db 331 VYSFSTDLKLVDTNKVQDYFKEFLQEFPR-LTHVSQQAIPVSATNAVENLNVLKKVKA- 388
QY 335 FQDNNLRNVMLPDGLEKLESEAFTGNGPGEHHNNQWLRTRTGPOLATEVYVNPDK 394
Db 389 ----NLNLVSP-----TQNFDFYFVNQHQKLEFGLEPNLTKOGL---E 428
QY 395 SLWRATPDMDYTKWLEDF-TY--QKNSVTGFSNKG--LOKVRRNKLEIPKQHNGITIT 449
Db 429 NLLISIKQSKNLKELRLNFYTYVAQETSRKQILKQATIKNLKRNKNQOBETPETKDETPS 488
QY 450 EIGDNAFRNTDFOSKTLRKYDLEEIKLPST---IRKIGAFAFQSNLKSFEAS--EDLEE 504
Db 489 E-STSGMKFDDHLSELTELEDF-SVNQATOETIYDSLHKILLIRSTNLKFKLSSYKMEMK 546
QY 505 IKGAF--MNRIGTL-DLKDKLLIKIGDAAFHINHIXAIVLPESV-----QEI--G 550
Db 547 SKMDTFLDILKVNLYETLNMLKRCVSVNISNP--HGNISYELINKDSTFYFKLTNLQHA 604
QY 551 RSAFRQNGALHLMFIGNKVKTIGEMAFLSNKLESV-----NLSEQKQQLTIEVQAFS 602
Db 605 KYTFQN---EFQFNMTVS---AKIESSSLSEDIDSLCKSIAASCKNLQNV----- 650
QY 603 DNALSEVVLPPNLQTIREFAKRNL---KEVKGSSTLSQITFNA-FDQNDGDKREGKKV 658
Db 651 -NIIASLILYPNQ---KNPFPNKPNLFFKQFELKNLENVSNICLQD----- 695
QY 659 VVRTHNNSHMLADGERFIELDPKLSSTMVDLEKVLKLTIEGLDYSTLROTTQ----- 709
Db 696 -----HILNSISEFLEXKKKIKAFILKRYVLLQY--LDYTKLFKTLQQLPELNQVY 745
QY 710 --TORREMNTAGKALLSKSNLRLQGEKOK-----FLQEAQFFLGRVDDKALA--KA 756
Db 746 INQOLEELTVSE---VHKQWENHQRKAFYEPCLCEFIKESQTLQOLIDFDQNTVSDDST 801
QY 757 EKALVTKKATKNGHLLERSINKAVLAYNSAIKKANVRLEKEELDLITDIVEKGGPLAQ 816
Db 802 KKLSEISESKYHHLRLNPSQ---SSSLIKSEN---EEIQELLKACDEKGVLVKA 851
QY 817 TMVQGVYLLKTPPLPEYYIGLNVYFD 843
Db 852 -----YVKEFPLCLP---TGTYYD 866

RESULT 13

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human

Query Match 3.6%; Score 192.5; DB 4; Length 2662;
Best Local Similarity 20.0%; Pred. No. 0.0002;
Matches 242; Conservative 170; Mismatches 433; Indels 365; Gaps 59;
Db 1387 QSLNMKEDNE---TTKIVSEMEQFKEKD--SALLRIE-IEMLGLSK--RQESHDE 1435
QY 869 PILNVE-DNEGHTLAVATLADYEGLYIKDILNSLDKIKAIROIPIAKYHRL-GIFQA 926
Db 1329 -RRLNNEKFQESQEEIKSLKRDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSKQE 1386

```

Qy 927 IRNAAAEADRLLPKTPKGYLNEVPNPKKKOMEKNLKPVDYKTPFENKALPNEKVGDRAA 986
 Db 1436 MKSVAKEKDDL----ORLQEVLOQHSQSDQLKENIK----EIVAKHLET---EELK 1479
 Qy 987 KGH---NINAETNNSAVTPIRSE-----QOLHKSQSDVNLPOT 1022
 Db 1480 VAHCCLKEQEETINELRVNLSEKETEISTIQKOLEAINDKLQNLQEIYERBQNLQI 1539
 Qy 1023 SSKNNFIVEI 1032
 Db 1540 SEVOENVNEI 1549

RESULT 14

US-09-107-532A-3959
 ; Sequence 3959, Application US/09107532A
 ; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESSEES: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 024354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107, 532A
 FILING DATE: 30-Jun-1998
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 60/085, 598
 FILING DATE: 14 May 1998
 FILING DATE: JULY 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3959:

SEQUENCE CHARACTERISTICS:

LENGTH: 1221 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1..1221
 SEQUENCE DESCRIPTION: SEQ ID NO: 3959:
 US-09-107-532A-3959

Query Match 3.5%; Score 189; DB 4; Length 1221;
 Best Local Similarity 19.0%; Pred. No. 0.00012; Indels 394; Gaps 61;
 Matches 228; Conservative 174; Mismatches 403; Indels 394; Gaps 61;

Qy 4 KHLKTLALAT-TIVSVVTYSQEVY--GLEREEESVKQEQTQSASEDDWFEEDNERKTNVS 59
 Db 14 KKLKLISIALFSLITMSTARQSVSAEGTSTKAVVEESGDKVAIDDKVKNEAEKATVQ 73
 Qy 60 KENS---TVDETVDLFDGNS--NNSSXTESV--VSDPKQ--VPKAKPENTQEASN 108

Qy 74 GADGIMIVVVDENDPEKLEAESPLKNOQENTGAVEVGVDERKQEKADKANEKVNFDI 133
 Db 109 SSNDASKVVEY---PK---QDTASKKETLETSTWEAKDFV--TRGDTLVGFSKSGINK 157
 Qy 134 DYNDISSLQSAALNDPKPFPLSEDELNEKLOKTKNTFESEALMKTR----GYFSQRIE 188
 Db 158 LSQTSHLV---LPSHAADGTQLTOVASFAFTPDKKTAIAEYTSRLGENG----KPS 206
 Db 189 LSNNLNDVPKRAKVSRAHALVSNRASQVSETRINPPNGYPWVYIIR--SNGSYTENYAI 246

Qy 207 RLIDQKETI-----DEGEIFNAYQLTKLIPNGYKSIG----- 240
 Db 241 ---QDAFVDNKNTAENVLPESLETISDY-----AFAHMSLKVQVQLPDLNLKV 284
 Db 307 WIATQMLVWENSN---EKPKSITVTPNWNGATKWNDLVARANAIRVRTKWDSTSRTV 362

Qy 285 --GELAFF----DNOI-----GGKLYLPRHLIKLAERAFKSNRQIOTVEFL--- 323
 Db 363 DEGDKATEKVNTYGDHNLHVKSVSQGGKATISNGTILTVDITGATGNKI-TVDMIKGSHP 421

Qy 324 -----GSKLKVIGEASFQDNRL-NVMLPDGLE--KIESE-----AF 357
 Db 422 SDAHPTTVWSNGSYQKLVTGSFYISTSVTNVKRGTVEVLKVDSETKKPLSGATFRFSY 481

Qy 358 TGNPGDEHYNQVVLRTGQNPHQULATE---NTYVNPDKSLWRATPDMDYTKWLEEDF 413
 Db 482 SGKTKD-----VTTDSNGKAKLTERLKGATVVK-----VKEIKAPN 517

Qy 414 TYQKN---SVTGFNSNKQGLOQKVRNRNKNLEIPKOHNGITITEIGD--NAFRNVD----- 461
 Db 518 GYQOLDSEFSFISITVKENQNVTTRNK----KSTGSVETEKIGDLDGLGLPNNVFTIYNS 572

Qy 462 QSKTLR-----KYLEEIKLPS--TIRKIGAFAFQSNNLK-SFEASEDLEIKEGA 509
 Db 573 DNKVVKDNLKIDSNGKXKVDLOFGKVIAYEVKQGVTGYDPDGKYNFETISQDTPFTK-- 629

Qy 510 FMMNRIGTLDDLKDKLJIGDAFHINHITYAIVLPESVQEIERSAFRQNGALHLMFIGNKV 569
 Db 630 -----PAKVVTNIVEVSISEKVSMDMNEQVLLNN----- 659

Qy 570 KTIGEMAFLSNKLESVNL--SEQQQLKTIIEVQAFSDNALSEVVLPPNQLTIREAFKRNH 627
 Db 660 WSADPNDYLQYDITTGNIKKSGGERNVQSFIEGFYDN-----KNDVSKSEVY---- 707

Qy 628 LKEVKGSSTLSQITFNAFDQNDGDKRFGKKVVRTHNShMLADGERFIIDPDKLSSSTMV 687
 Db 708 -----VGSTNTVT----NSFDINNDT-----NNGKVIAKAKTTSVLNTNDFYNSY 747

Qy 688 DLEKVLKLIETGLDYSLTRQTTQFREMTAGKALLSKSMLROGEKQKFLQEAQFFLGRV 747
 Db 748 NLRITMKI-----KKSTLENTEKQKNL----- 769

Qy 748 DLKAIKAKAELVTKKATRKHILLERSINKAVLAVNNSAIKKANVRL-EKBLDILTL 806
 Db 770 ---TIVKNSKVTVGSKSKSA---SSNDVETVLYS---RKVTINHIDEKDKHLLQD 817

Qy 807 VEGK--GLAQATMVQGVVLLKTPPLPEYYIGLNVYFDKSGKLIYALDMSDTIGEGOKD 864
 Db 818 IDYKYDGETYE-----YKPRTDL-----FDKDGNNY---KSDVTHKGKID 854

Qy 865 AYGNPILINVDEDNEGTYHTLAVATLADYEGLYIKDILNSSLDKTKAIROIPLAKYHRLGIF 924
 Db 855 --GKDIVL---NTPYHT-----PVLSVNDRI---QIDTAR----- 882

Qy 925 QAIRNAAAEADRLLPKTPKGYL-NEVPNPKKKOMEKNLKPVDYKTPFENKALPNEKVGDG 983
 Db 883 -AVKN-----GYLPTKIELSKKEYEKELEKIVFKVKITD--IDNKKVYD 925

Qy 984 RAAKGHNIAETNNSAVTPIRSEQQLHKSQSDVNLPOTSKNNFIVE----ILGYVS 1037

Db 926 ENFKYKEFNHMSLDLPTEYLTKKKKINY-S-VDILLVDPDKNKFETETKDLRTIGTS 983

RESULT 15

US-08-592-126-148

; Sequence 148, Application US/08592126

; Patent No. 5821091

; GENERAL INFORMATION:

; APPLICANT: Gregory Dolganov

; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

; NUMBER OF SEQUENCES: 151

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dehlinger & Associates

; STREET: 350 Cambridge Avenue, Suite 250

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #11.0, Version #11.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592,126

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:

LENGTH: 1312 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54

; US-08-592-126-148

Query Match 3.5%; Score 188; DB 2; Length 1312;

Best Local Similarity 19.2%; Pred. No. 0.00015; Mismatches 446; Indels 296; Gaps 50; Matches 220; Conservative 185; Mismatches 446; Indels 296; Gaps 50;

QY 18 VVTYQSEVYGLE-REESVKQEQTQASADEDDWFEEDNERKTVSKENSTVDETVSDLF--S 74

Db 241 VPKSYENELDPLKNRKLIEHNLSKIMKLDNEIKALDSRKQMEKUNSELEEKMEKVFGQT 300

QY 75 DGNSSNNSSKTESVSDPKQVKPKAEPVQEQEASNSNDASKVEVPQDQTASKKETLETST 134

Db 301 DEQNLNDLYHNHQTIV--REKERKUVCHRELEKLNKESRLLNQEKSELLVQGRLOQA 357

QY 944 GYLNEVNPYRKKOMEKVLKPVDYK-----TPIFNKALPNEKVGDRA-AKGHNIN 992

Db 1080 GYEEELIHTFKELREPQFRDAEKKYREMMIVMVRTTELVNKDLDIYKTLQAIMKEHSMK 1139

QY 993 AETNNISVAVTPIRSE--QQLH---KSQSDWNLPTSSKNNFIYETL--GYVSLCLLF 1042

Db 1140 MEELINKLIRDLWRSTYRGQDIEYIEIRSDADENVSA SDKRNNYNTVVMLKGDTALDMRG 1199

QY 1043 LVTAGKK 1049

Db 1200 RCSAGQK 1206

QY 135 WEAKDFVTRGDTLVLGFSKSGINKLSQLTSHNLVPLSHAADGTLQTOVAEFTPDKTAIE 194

Db 358 DRHOEHTRARDSLI-----QSLATOL-ELDGFERGPFSERQIKN 395

QY 195 YTSRSLGENGKPSRLIDQKEIIDEGEIENFAAYQLTKLTPNGYKSIGQDAFVNKNIAEVN 254

Db 396 FHKLVRE-----QEGEAKTANQL-----MNDFAEKETLKKQ 428

QY 255 LPE-----SLETISDYAFAHMSLKVQVQLPDLNKGELAFFDNQIGGKYLPRHLIKLA 308

Db 429 IDEIRDKKTGIGRIBELKSEILSKKQNEI-KNVKY--ELOQLEGSSDRILELDOELIKE 485

QY 309 ERAFKSNRIQTVEFLGSKLKV-----GEASFQDNLRLNVMPLPDGLEKIESEAFTGNPGDEH 365

Db 486 RELSKAENSNVETL--KMEVLSQNEKAQDLRTR-----KLDQEM-----EQ 527

QY 366 YNNQVVLRTGQMPHOLATENTVNPDKSIN--RATPMDY--TKWLEEDFTYQKNSVT 421

Db 528 LNHHTTTRTQMEMLTKDKADKDEQIRKIKSRHSDELTSLIGYFPNKKQLEDWLHHSKEI 587

QY 422 GFS-----NKGLOKVRNK---NLEIPKQHNGITITEIGDNAF--RNVDFOSKTLR 468

Db 588 NOTRDRLAKNLKELASSEQNKNHINNELLKREEQLSSYE--DKLFDVCGSQFES--- 640

QY 469 YDLI---EEIKLPSTIRKI--GAFAPQSNL-----KSFEASEDLEEIKE 507

Db 641 -DLDRIKEEIEKSSKQRAMLAGATAVYSQFITQLTDENQSCCPVCVQRFQTEAELQEV-- 697

QY 508 GAFMNMRIGTLDLKDQLIKIGDAFAFHINHYATIVLPESWQETGRSAFRQNGAHLHM--- 563

Db 698 -----ISDIQSKRLAPDK-----LKSTESELKKERDRDEMILGLVPMRQ 737

QY 564 -FIGNKVKTIGEMAFLSNKLESVNLSQEQKQKTIIEVQAFSDNALESVULPPNLQTRREA 622

Db 738 SIDLKEKEIPE--LRNKLQVNVDIQRKNDIEQ---ETLGLTIMP-----EES 784

QY 623 FK-----RNHILKEVKGSTSLSQITNAFDQNDGDKRFGKVVVRTHN----- 664

Db 785 AKVCLTDVTIMERFQMLKIDVERKIAQQAQKLOGIDLDRTVQVNQEKQEKQHKLDTVSS 844

QY 665 ---NSHMLADGERFTIDDKLSSTMVDLEKVLTIEGLDYST--LRQTTQTFREMT 717

Db 845 KIELNRKLQDQEQI---QHLKSTTNEIKS-----EKLOISTNLQRROQLEBQTVELST 896

QY 718 AGKALLISKSINLROGEKQ-----KFLQEAQFFLGRVLDKATAKEAKALVTKATK 768

Db 897 EVQSLYRE--IKDAKEQVSPLETLEKFQOEKEELINKKNTSNKIAQ-DKLDIKEKVKN 953

QY 769 GHLLERSINKAVLAVNNSAIKKANVKRLEKEIDDLITDVLVEKGKPLAQ--ATMVQGVYLLK 826

Db 954 IHGIMKDIEI-YIQDGKDDYKKQKETELNKVIAQLSECEKHKEKINEDMRLMQRQDIDTQK 1012

QY 827 TPLPLPEYVYGLNVYFDKSGKLUYALDMSDTIGEGQKDAGNPLINVDEDNEGHTLAVA 886

Db 1013 -----IQERWLQDNLTLRKRNE-----ELKEVEERKQH---LK 1043

QY 887 TLADYEGLYK--DILNSSLDKIKAIROQIPLAKYHRLGIFQAIRNAAEADRILLPKTP 943

Db 1044 EMGQMOVLOMKSEHOKLEENIDNIKRHNHLAIGR-----OK 1079

Search completed: April 27, 2004, 11:18:08

Job time : 28 secB